ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ

A Systematic Review and Meta-Analysis of the Association Between Advanced Dental Disease and Severe Mental Illness Using R software.

Παναγιώτα Σάββα Αληφραγκή

ΕΡΓΑΣΙΑ

Που υποβλήθηκε στο Τμήμα Στατιστικής
του Οικονομικού Πανεπιστημίου Αθηνών
ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος
Συμπληρωματικής Ειδίκευσης στη Στατιστική
Μερικής Παρακολούθησης (Part-time)

Αθήνα
Νοέμβριος 2017

ΑΦΙΕΡΩΣΗ

Στον καθηγητή μου, κύριο Βασδέκη Βασίλειο.

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ τον καθηγητή κ. Βασδέκη για το σημαντικό έργο που επιτελεί στο τμήμα μας, την πίστη και αφοσίωσή του στους φοιτητές, την ανάθεση της εργασίας, τις λεπτομερείς διορθώσεις του, την καθοδήγησή του, την υπομονή του και για την άψογη συνεργασία μας.

Ευχαριστώ την οικογένεια μου και τους φίλους μου που πίστεψαν σε εμένα και με στήριξαν σε όλη την πορεία μου, στην ολοκλήρωση των μεταπτυχιακών μου σπουδών και της διπλωματικής μου εργασίας.

ΒΙΟΓΡΑΦΙΚΟ ΣΗΜΕΙΩΜΑ

Ονομάζομαι Αληφραγκή Παναγιώτα και γεννήθηκα στον Πειραιά στις 20 Σεπτεμβρίου του 1989. Πήγα σχολείο στον Πειραιά και το 2017 εισήχθηκα στην Οδοντιατρική Σχολή του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης. Το 2015 έγινα δεκτή στο μεταπτυχιακό πρόγραμμα σπουδών Εφαρμοσμένης Στατιστικής με κατεύθυνση Ιατρική μερικής φοίτησης στο Οικονομικό Πανεπιστήμιο Αθηνών, το οποίο και ολοκλήρωσα το 2017. Συνεχίζω τις σπουδές μου με σκοπό το διδακτορικό στην Ιατρική Σχολή Αθηνών στο τμήμα Επιδημιολογίας και Βιοστατιστικής. Όλα αυτά τα χρόνια εργάστηκα ως οδοντίατρος σε ιδιωτικά ιατρεία αλλά και εθελοντικά. Ασχολήθηκα επίσης και με το σχεδιασμό και την ανάλυση δεδομένων. Μιλώ Αγγλικά και Γαλλικά σε επίπεδο C2 και B1 αντίστοιχα, έχω άριστη γνώση ηλεκτρονικών υπολογιστών καθώς και πολύ καλή γνώση στατιστικών πακέτων ανάλυσης την οποία και απέκτησα κατά τη φοίτησή μου στο μεταπτυχιακό πρόγραμμα σπουδών. Έχω συμμετάσχει στην συγγραφή δύο προς δημοσίευση ερευνών, οι οποίες έχουν παρουσιασθεί σε ευρωπαϊκά συνέδρια.

ABSTRACT

PANAGIOTA ALIFRAGKI

A Systematic Review and Meta-Analysis of the Association Between Advanced Dental Disease and Severe Mental Illness using R software.

November 2017

The poor physical health faced by people with mental illness has been the subject of growing attention, but there has been less focus on the issue of oral health even though it is an important part of physical health. Psychiatric patients have increased co morbid physical illness. There is less information concerning dental disease in this population in spite of risk factors including diet and psychotropic side-effects (such as xerostomia). This paper is based on Kisely et. al. (2011) and its aim was to compare the oral health of people with severe mental illness with that of the general population. Another aim of the thesis is to provide a real data comparison between three fixed effects methods for the calculation of the fixed effects estimate (Mantel-haenszel, Peto and Inverse variance).

A systematic search for studies from the past 20 years was conducted using Medline, PsycINFO, Embase and article bibliographies. Papers were independently assessed. The primary outcome using the package R and the library "meta" and "metafor" for the analysis of the data was: total tooth loss (edentulousness), the end-stage of both untreated caries and periodontal disease, and dental decay through standardized measures: the mean number of decayed, missing and filled teeth (DMFT) or surfaces (DMFS). For studies lacking a control group controls of similar ages from a community survey within 10 years of the study were used.

As result 21 papers were identified of which 14 had sufficient data (n = 2784 psychiatric patients) and suitable controls (n = 31 084) for a fixed effect and a random effects meta-analysis. People with severe mental illness had 3.4 times the odds of having lost all their teeth than the general community (95% CI 1.6–7.2). They also had significantly higher scores for DMFT (mean difference 6.2, 95% CI 0.6–11.8) and

DMFS (mean difference 14.6, 95% CI 4.1–25.1). Fluoridated water reduced the gap in oral health between psychiatric patients and the general population. Psychiatric patients have not shared in the improving oral health of the general population. Management should include oral health assessment using standard checklists that can be completed by non-dental personnel. Interventions include oral hygiene and management of xerostomia.

ΠΕΡΙΛΗΨΗ

ΠΑΝΑΓΙΩΤΑ ΑΛΗΦΡΑΓΚΗ

Μια συστηματική αναθεώρηση και μετα-ανάλυση της σύνδεσης μεταξύ της προχωρημένης οδοντικής νόσου και της σοβαρής ψυχικής ασθένειας, χρησιμοποιώντας τη γλώσσα R.

Νοέμβριος 2017

Η κακή σωματική υγεία που αντιμετωπίζουν τα άτομα με ψυχικές ασθένειες έχει αποτελέσει αντικείμενο αυξανόμενης προσοχής, αλλά έχει επικεντρωθεί λιγότερο στο θέμα της στοματικής υγείας, παρόλο που αποτελεί σημαντικό μέρος της σωματικής υγείας. Οι ψυχιατρικοί ασθενείς έχουν αυξημένη συνυπάρχουσα σωματική ασθένεια. Υπάρχουν λιγότερες πληροφορίες σχετικά με τις οδοντικές ασθένειες σε αυτόν τον πληθυσμό, παρά τους παράγοντες κινδύνου, συμπεριλαμβανομένων των διατροφικών και ψυχοτρόπων παρενεργειών (όπως η ξηροστομία). Αυτό το έγγραφο βασίζεται στη δημοσίευση των Kisely et. al. (2011) και στόχος του ήταν να συγκρίνει την στοματική υγεία των ατόμων με σοβαρές ψυχικές ασθένειες με αυτά του γενικού πληθυσμού. Ένας άλλος στόχος της εργασίας είναι να υπάρξει πραγματική σύγκριση δεδομένων μεταξύ τριών μεθόδων για τον υπολογισμό της εκτίμησης των αποτελεσμάτων (Mantel-haenszel, Peto και Inverse variance). Μια συστηματική αναζήτηση για μελέτες από τα τελευταία 20 χρόνια διεξήχθη χρησιμοποιώντας Medline, PsycINFO, Embase και βιβλιογραφίες άρθρων. Τα έγγραφα αξιολογήθηκαν ανεξάρτητα. Το πρωτεύον αποτέλεσμα με τη χρήση της γλώσσας R και της βιβλιοθήκης «meta» και «metafor» για την ανάλυση των δεδομένων ήταν: ολική απώλεια των δοντιών (edentulousness), το τελικό στάδιο τόσο της τερηδόνας όσο και της περιοδοντικής νόσου και της οδοντικής αποσύνθεσης μέσω τυποποιημένων μέτρων : ο μέσος αριθμός των τερηδονισμένων, ελλειπόντων και εμφραγμένων δοντιών (DMFT) ή επιφανειών (DMFS). Για μελέτες που δεν είχαν ομάδα ελέγχου, χρησιμοποιήθηκαν έλεγχοι παρόμοιων ηλικιών από μια κοινοτική έρευνα εντός 10 ετών από τη μελέτη. Ως αποτέλεσμα, προσδιορίστηκαν 21 έγγραφα από τα οποία 14 είχαν επαρκή δεδομένα (n = 2784 ψυχιατρικούς ασθενείς) και κατάλληλους μάρτυρες (n = 31 084)

για σταθερό αποτέλεσμα και μετα-ανάλυση τυχαίων αποτελεσμάτων. Τα άτομα με σοβαρή διανοητική ασθένεια είχαν 3,4 φορές περισσότερες πιθανότητες να χάσουν όλα τα δόντια τους από τη γενική κοινότητα (95% CI 1,6-7,2). Επίσης, είχαν σημαντικά υψηλότερες βαθμολογίες για DMFT (μέση διαφορά 6.2, 95% CI 0.6-11.8) και DMFS (μέση διαφορά 14.6, 95% CI 4.1-25.1). Το φθοριωμένο νερό μείωσε το κενό στην στοματική υγεία μεταξύ των ψυχιατρικών ασθενών και του γενικού πληθυσμού. Οι ψυχιατρικοί ασθενείς δεν έχουν συμμετάσχει στη βελτίωση της στοματικής υγείας του γενικού πληθυσμού. Η διαχείριση θα πρέπει να περιλαμβάνει την εκτίμηση της στοματικής υγείας χρησιμοποιώντας τυποποιημένους καταλόγους ελέγχου που μπορούν να συμπληρωθούν από μη οδοντιατρικό προσωπικό. Οι παρεμβάσεις περιλαμβάνουν την στοματική υγιεινή και τη διαχείριση της ξηροστομίας.

ΚΑΤΑΛΟΓΟΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Σελίδα

CHAPTER 3: Statistical Analysis and Methods

CHAPTER 4: Conclusion

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

Πίνακας Σελίδα

ΚΑΤΑΛΟΓΟΣ ΓΡΑΦΗΜΑΤΩΝ

Γράφημα Σελίδα

CHAPTER 1: INTRODUCTION

This paper is based on Kisely et al., (2011). The aim of this research paper is to compare the oral health of people with severe mental illness with that of the general population.

It is well known that individuals with severe mental illness (SMI) have high rates of physical ill-health including diabetes, cardiovascular disease, chronic lung disease, and cancer (Lawrence et al., 2000). This, in turn, is associated with increased mortality from preventable physical disease so that people with schizophrenia die 15 to 20 years earlier than the general population. Oral health is an important part of physical health. (Mirza et al., 2003). Historically, there has been less attention to the issue of oral health, although it is also an important part of physical health and linked to systemic diseases such as coronary heart disease, stroke, diabetes, and respiratory disease. Oral health also affects eating, speech and other social and psychological areas of life. People with severe mental illness are susceptible to oral disease for a number of reasons: these include amotivation, poor oral hygiene, fear, specific dental phobia, dental costs, difficulty in accessing healthcare facilities and the side-effects of psychiatric drugs such as dry mouth (xerostomia) (Bardow et al., 2001).

The two most common diseases that affect oral health are dental caries (tooth decay) and periodontal disease. Dental caries occurs through the demineralization and subsequent proteolysis of the hard tooth structure (enamel and dentine) from a build-up of dental plaque which microorganisms colonize. If plaque is not removed, and there is frequent intake of readily fermentable carbohydrates in the diet, irreversible cavitation can occur. This will normally require restoration or extraction of the tooth if the dental pulp has become infected. Periodontal disease usually begins with gingivitis – inflammation of the gingival tissues (gums). This, too, is caused by longstanding accumulation of dental plaque in contact with the soft tissues. In patients who harbor particularly pathogenic microflora, or whose host response to these micro-organisms is ineffective, inflammation spreads to the periodontal ligament with destruction of connective tissues and surrounding (alveolar) bone. Signs of periodontal disease include bleeding gums and pockets where the gingivae have become detached from the teeth. In more advanced disease there is exposure of tooth

roots and mobility of teeth (Pihlstrom et al., 2005). These symptoms and signs are often associated with halitosis (bad breath). The end-stage of both untreated dental caries and periodontal disease is tooth loss, which can involve the whole dentition (edentulousness). In an earlier meta-analysis, the authors reported significantly higher levels of edentulousness in patients with SMIs such as dementia, schizophrenia, bipolar affective disorder, and other affective disorders. By contrast, the effect on other measures of oral health, such as dental decay, was inconclusive. This was possibly because of the low number of studies that could be incorporated into meta-analyses.

The research team of the paper therefore focused on this condition as an indicator of both dental caries and periodontal disease in people with severe mental illness. To our knowledge, this is the first systematic review and meta-analysis of this topic in people with severe mental illness. The effect of water fluoride levels on differences in oral health between people with and without severe mental illness was also considered. The aim, therefore, was to compare the prevalence of edentulousness in people with severe mental illness with that in the general population. Levels of dental decay were also compared. A systematic search for studies from the past 20 years was conducted using Medline, PsycINFO, Embase and article bibliographies was used for the paper that we based our analysis using software R and the library "meta" for the analysis of the data. Papers were independently assessed. The primary outcome was total tooth loss (edentulousness), the end-stage of both untreated caries and periodontal disease. Dental decay was also assessed through standardized measures: the mean number of decayed, missing and filled teeth (DMFT) or surfaces (DMFS). For studies lacking a control group we used controls of similar ages from a community survey within 10 years of the study.

The authors identified 21 papers of which 14 had sufficient data (n = 2784 psychiatric patients) and suitable controls (n = 31 084) for a random effects meta-analysis. People with severe mental illness had 3.4 times the odds of having lost all their teeth than the general community (95% CI 1.6–7.2). They also had significantly higher scores for DMFT (mean difference 6.2, 95% CI 0.6–11.8) and DMFS (mean difference 14.6, 95% CI 4.1–25.1). Fluoridated water reduced the gap in oral health between psychiatric patients and the general population. Psychiatric patients have not shared in the improving oral health of the general population. Management should include oral health assessment using standard checklists that can be completed by

non-dental personnel. Interventions include oral hygiene and management of xerostomia.

However, it is good to define what systematic review and meta-analysis is and how these procedures can work. S. Gopalakrishnan and P. Ganeshkumar present all the necessary theory and they present the basic steps for this analysis.

Systematic review

A systematic review is a summary of the medical literature that uses explicit and reproducible methods to systematically search, critically appraise, and synthesize on a specific issue. It synthesizes the results of multiple primary studies related to each other by using strategies that reduce biases and random errors. To this end, systematic reviews may or may not include a statistical synthesis called meta-analysis, depending on whether the studies are similar enough so that combining their results is meaningful. Systematic reviews are often called overviews.

The evidence-based practitioner, David Sackett, defines the following terminologies.

- Review: The general term for all attempts to synthesize the results and conclusions of two or more publications on a given topic.
- Overview: When a review strives to comprehensively identify and track down all the literature on a given topic (also called "systematic literature review").
- Meta-analysis: A specific statistical strategy for assembling the results of several studies into a single estimate.

Systematic reviews adhere to a strict scientific design based on explicit, pre-specified, and reproducible methods. Because of this, when carried out well, they provide reliable estimates about the effects of interventions so that conclusions are defensible. Systematic reviews can also demonstrate where knowledge is lacking. This can then be used to guide future research. Systematic reviews are usually carried out in the areas of clinical tests (diagnostic, screening, and prognostic), public health interventions, adverse (harm) effects, economic (cost) evaluations, and how and why interventions work.

Meta-analysis

A meta-analysis is the combination of data from several independent primary studies that address the same question to produce a single estimate like the effect of treatment or risk factor. It is the statistical analysis of a large collection of analysis and results from individual studies for the purpose of integrating the findings.The term meta-analysis has been used to denote the full range of quantitative methods for research reviews. Meta-analyses are studies of studies.Meta-analysis provides a logical framework to a research review where similar measures from comparable studies are listed systematically and the available effect measures are combined wherever possible.

The fundamental rationale of meta-analysis is that it reduces the quantity of data by summarizing data from multiple resources and helps to plan research as well as to frame guidelines. It also helps to make efficient use of existing data, ensuring generalizability, helping to check consistency of relationships, explaining data inconsistency, and quantifies the data. It helps to improve the precision in estimating the risk by using explicit methods.

Therefore, "systematic review" will refer to the entire process of collecting, reviewing, and presenting all available evidence, while the term "meta-analysis" will refer to the statistical technique involved in extracting and combining data to produce a summary result.

Following are the six fundamental essential steps while doing systematic review and meta-analysis.

Define the question

This is the most important part of systematic reviews/meta-analysis. The research question for the systematic reviews may be related to a major public health problem or a controversial clinical situation which requires acceptable intervention as a possible solution to the present healthcare need of the community. This step is most important since the remaining steps will be based on this.

Reviewing the literature

This can be done by going through scientific resources such as electronic database, controlled clinical trials registers, other biomedical databases, non-English literatures, "gray literatures" (thesis, internal reports, non–peer-reviewed journals, pharmaceutical industry files), references listed in primary sources, raw data from published trials and other unpublished sources known to experts in the field. Among the available electronic scientific database, the popular ones are PUBMED, MEDLINE, and EMBASE.

Sift the studies to select relevant ones

To select the relevant studies from the searches, we need to sift through the studies thus identified. The first sift is pre-screening, i.e., to decide which studies to retrieve in full, and the second sift is selection which is to look again at these studies and decide which are to be included in the review. The next step is selecting the eligible studies based on similar study designs, year of publication, language, choice among multiple articles, sample size or follow-up issues, similarity of exposure, and or treatment and completeness of information.

It is necessary to ensure that the sifting includes all relevant studies like the unpublished studies (desk drawer problem), studies which came with negative conclusions or were published in non-English journals, and studies with small sample size.

Assess the quality of studies

The steps undertaken in evaluating the study quality are early definition of study quality and criteria, setting up a good scoring system, developing a standard form for assessment, calculating quality for each study, and finally using this for sensitivity analysis.

For example, the quality of a randomized controlled trial can be assessed by finding out the answers to the following questions:

1. Was the assignment to the treatment groups really random?
2. Was the treatment allocation concealed?

3. Were the groups similar at baseline in terms of prognostic factors?
4. Were the eligibility criteria specified?
5. Were the assessors, the care provider, and the patient blinded?
6. Were the point estimates and measure of variability presented for the primary outcome measure?
7. Did the analyses include intention-to-treat analysis?

Calculate the outcome measures of each study and combine them

We need a standard measure of outcome which can be applied to each study on the basis of its effect size. Based on their type of outcome, following are the measures of outcome: Studies with binary outcomes (cured/not cured) have odds ratio, risk ratio; studies with continuous outcomes (blood pressure) have means, difference in means, standardized difference in means (effect sizes); and survival or time-to-event data have hazard ratios.

Combining studies

Homogeneity of different studies can be estimated at a glance from a forest plot (explained below). For example, if the lower confidence interval of every trial is below the upper of all the others, i.e., the lines all overlap to some extent, then the trials are homogeneous. If some lines do not overlap at all, these trials may be said to be heterogeneous.

The definitive test for assessing the heterogeneity of studies is a variant of Chi-square test (Mantel–Haenszel test). The final step is calculating the common estimate and its confidence interval with the original data or with the summary statistics from all the studies. The best estimate of treatment effect can be derived from the weighted summary statistics of all studies which will be based on weighting to sample size, standard errors, and other summary statistics. Log scale is used to combine the data to estimate the weighting.

Interpret results: Graph

The results of a meta-analysis are usually presented as a graph called forest plot because the typical forest plots appear as forest of lines. It provides a simple visual presentation of individual studies that went into the meta-analysis at a glance. It

shows the variation between the studies and an estimate of the overall result of all the studies together.

Subgroup analysis

Subgroup analysis looks at the results of different subgroups of trials, e.g., by considering trials on adults and children separately. This should be planned at the protocol stage itself which is based on good scientific reasoning and is to be kept to a minimum.

CHAPTER 2:
METHODS OF META-ANALYSIS

2.1 INTRODUCTION TO META-ANALYSIS

Michael Borenstein in his book Introduction to Meta-analysis (2009) presents all the theory about the history, the procedure and the aim of Meta-analysis. Prior to the 1990s, the task of combining data from multiple studies had been primarily the purview of the narrative review. An expert in a given field would read the studies that addressed a question, summarize the findings, and then arrive at a conclusion – for example, that the treatment in question was, or was not, effective. However, this approach suffers from some important limitations.

One limitation is the subjectivity inherent in this approach, coupled with the lack of transparency. For example, different reviewers might use different criteria for deciding which studies to include in the review. Once a set of studies has been selected, one reviewer might give more credence to larger studies, while another gives more credence to 'quality' studies and yet another assigns a comparable weight to all studies. One reviewer may require a substantial body of evidence before concluding that a treatment is effective, while another uses a lower threshold. In fact, there are examples in the literature where two narrative reviews come to opposite conclusions, with one reporting that a treatment is effective while the other reports that it is not. As a rule, the narrative reviewer will not articulate (and may not even be fully aware of) the decision-making process used to synthesize the data and arrive at a conclusion.

A second limitation of narrative reviews is that they become less useful as more information becomes available. The thought process required for a synthesis requires the reviewer to capture the finding reported in each study, to assign an appropriate weight to that finding, and then to synthesize these findings across all studies in the synthesis. While a reviewer may be able to synthesize data from a few studies in their head, the process becomes difficult and eventually untenable as the number of studies increases. This is true even when the treatment effect (or effect size) is consistent from study to study. Often, however, the treatment effect will vary as a function of study level covariates, such as the patient population, the dose of medication, the outcome variable, and other factors. In these cases, a proper synthesis

requires that the researcher be able to understand how the treatment effect varies as a function of these variables, and the narrative review is poorly equipped to address these kinds of issues.

For these reasons, beginning in the mid 1980s and taking root in the 1990s, researchers in many fields have been moving away from the narrative review, and adopting systematic reviews and meta-analysis.

For systematic reviews, a clear set of rules is used to search for studies, and then to determine which studies will be included in or excluded from the analysis. Since there is an element of subjectivity in setting these criteria, as well as in the conclusions drawn from the meta-analysis, we cannot say that the systematic review is entirely objective. However, because all of the decisions are specified clearly, the mechanisms are transparent.

A key element in most systematic reviews is the statistical synthesis of the data, or the meta-analysis. Unlike the narrative review, where reviewers implicitly assign some level of importance to each study, in meta-analysis the weights assigned to each study are based on mathematical criteria that are specified in advance. While the reviewers and readers may still differ on the substantive meaning of the results (as they might for a primary study), the statistical analysis provides a transparent, objective, and replicable framework for this discussion.

The formulas used in meta-analysis are extensions of formulas used in primary studies, and are used to address similar kinds of questions to those addressed in primary studies. In primary studies we would typically report a mean and standard deviation for the subjects. If appropriate, we might also use analysis of variance or multiple regression to determine if (and how) subject scores were related to various factors. Similarly, in a meta-analysis, we might report a mean and standard deviation for the treatment effect. And, if appropriate, we would also use procedures analogous to analysis of variance or multiple regression to assess the relationship between the effect and study-level covariates.

Meta-analyses are conducted for a variety of reasons, not only to synthesize evidence on the effects of interventions or to support evidence-based policy or practice. The purpose of the meta-analysis, or more generally, the purpose of any research synthesis has implications for when it should be performed, what model should be used to analyze the data, what sensitivity analyses should be undertaken, and how the results should be interpreted. Losing sight of the fact that meta-analysis

is a tool with multiple applications causes confusion and leads to pointless discussions about what is the right way to perform a research synthesis, when there is no single right way. It all depends on the purpose of the synthesis, and the data that are available.

Meta-analysis was defined by Glass (1976) to be 'the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings'. Although Glass was involved in social science research, the term 'meta-analysis' has been adopted within other disciplines and has proved particularly popular in clinical research. Some of the techniques of meta-analysis have been in use for far longer. Pearson (1904) applied a method for summarizing correlation coefficients from studies of typhoid vaccination, Tippet (1931) and Fisher (1932) presented methods for combining *p*-values, and Yates and Cochran (1938) considered the combination of estimates from different agricultural experiments. However, the introduction of a name for this collection of techniques appears to have led to an upsurge in development and application.

Systematic reviews and meta-analyses are used to synthesize the available evidence for a given question to inform policy, as in the examples cited above from medicine, social science, business, ecology, and other fields. While this is probably the most common use of the methodology, meta-analysis can also play an important role in other parts of the research process. Systematic reviews and meta-analyses can play a role in designing new research. As a first step, they can help determine whether the planned study is necessary.

It may be possible to find the required information by synthesizing data from prior studies, and in this case, the research should not be performed. Iain Chalmers (2007) made this point in an article entitled The lethal consequences of failing to make use of all relevant evidence about the effects of medical treatments: the need for systematic reviews.

In the medical world, the upsurge began in the 1980s.Some of the key medical questions answered by meta-analyses at this time concerned the treatment of heart disease and cancer. For example, Yusuf *et al*.(1985) concluded that long-term beta blockade following discharge from the coronary care unit after amyocardial infarction reduced mortality, and the Early Breast Cancer Trialists' Collaborative Group (1988) showed that tamoxifen reduced mortality in women over 50 with early breast cancer .By the 1990s published meta-analyses were ubiquitous.

The rapid increase in the number of meta-analyses being conducted during the last decade is mainly due to a greater emphasis on evidence-based medicine and the need for reliable summaries of the vast and expanding volume of clinical research. Evidence-based medicine has been defined as 'integrating individual clinical expertise with the best available external clinical evidence from systematic research' (Sackett *et al*., 1997). A systematic review of the relevant external evidence provides a framework for the integration of the research, and meta-analysis offers a quantitative summary of the results. In many cases a systematic review will include a meta-analysis, although there are some situations when this will be impossible due to lack of data or inadvisable due to unexplained inconsistencies between studies.

In the event that the new study is needed, the meta-analysis may be useful in helping to design that study. For example, the meta-analysis may show that in the prior studies one outcome index had proven to be more sensitive than others, or that a specific mode of administration had proven to be more effective than others, and should be used in the planned study as well. For these reasons, various government agencies, including institutes of health in various countries, have been encouraging (or requiring) researchers to conduct a meta-analysis of existing research prior to undertaking new funded studies. The systematic review can also play a role in the publication of any new primary study. In the introductory section of the publication, a systematic review can help to place the new study in context by describing what we knew before, and what we hoped to learn from the new study. In the discussion section of the publication, a systematic review allows us to address not only the information provided by the new study, but the body of evidence as enhanced by the new study. Iain Chalmers and Michael Clarke (1998) see this approach as a way to avoid studies being reported without context, which they refer to as 'Islands in Search of Continents'. Systematic
reviews would provide this context in a more rigorous and transparent manner than the narrative reviews that are typically used for this purpose.

In accordance with ICH E9, meta-analysis is understood to be a formal evaluation of the quantitative evidence from two or more trials bearing on the same question. The guidelines indicate that meta-analysis techniques provide a useful means of summarizing overall efficacy results of a drug application and of analyzing less frequent outcomes in the overall safety evaluation. However, there is a warning that confirmation of efficacy from a meta-analysis only will not usually be accepted

as a substitute for confirmation of efficacy from individual trials. Certainly the magnitude of the treatment effect is likely to be an important factor in regulatory decision-making. If the treatment effect is smaller than anticipated, then statistical significance may not be reached in the individual trials. Even if statistical significance is reached in the meta-analysis, the magnitude of the treatment effect may not be *clinically* significant, and thus be considered insufficient for approval. Fisher (1999) considered the two conditions under which one large trial might substitute for the two controlled trials usually required by the Food and Drug Administration (FDA) in the USA. The first relates to the strength of evidence for demonstrating efficacy. He showed that if the evidence required from the two controlled trials is that they should each be statistically significant at the two-sided 5% significance level, then the same strength of evidence is obtained from one large trial if it is statistically significant at the two-sided 0.125% level. The same type of argument could be applied to combining trials in a meta-analysis.

It would seem reasonable to set a more stringent level of statistical significance corresponding to proof of efficacy in a meta-analysis than in the individual trials. The second condition discussed by Fisher relates to evidence of replicability, and he proposes criteria which need to be met by the one large trial. A meta-analysis will always involve at least two trials, and it will be important to assess the consistency of the results from the individual trials. The extent of any inconsistencies amongst the trials will be influential in the choice of model for the meta-analysis and in the decision whether to present an overall estimate**.**

WE USE META-ANALYSIS:

• To provide a more precise estimate of the overall treatment effects.

• To evaluate whether overall positive results are also seen in pre-specified subgroups of patients.

• To evaluate an additional efficacy outcome that requires more power than the individual trials can provide.

• To evaluate safety in a subgroup of patients, or a rare adverse event in all patients.

• To improve the estimation of the dose-response relationship.

• To evaluate apparently conflicting study results.

2.2 TREATMENT EFFECTS AND EFFECT SIZES

Borenstein (2009) and references within made an extensive review of the subject and presented all necessary theory. The terms treatment effects and effect sizes are used in different ways by different people. Meta-analyses in medicine often refer to the effect size as a treatment effect, and this term is sometimes assumed to refer to odds ratios, risk ratios, or risk differences, which are common in meta-analyses that deal with medical interventions. Similarly, meta-analyses in the social sciences often refer to the effect size simply as an effect size and this term is sometimes assumed to refer to standardized mean differences or to correlations, which are common in social science meta-analyses.

In fact, though, both the terms effect size and treatment effect can refer to any of these indices, and the distinction between these terms lies not in the index itself but rather in the nature of the study. The term effect size is appropriate when the index is used to quantify the relationship between two variables or a difference between two groups. By contrast, the term treatment effect is appropriate only for an index used to quantify the impact of a deliberate intervention. Thus, the difference between males and females could be called an effect size only, while the difference between treated and control groups could be called either an effect size or a treatment effect.

While most meta-analyses focus on relationships between variables, some have the goal of estimating a mean or risk or rate in a single population. For example, a meta-analysis might be used to combine several estimates for the prevalence of Lyme disease in Wabash or the mean SAT score for students in Utah. In these cases the index is clearly not a treatment effect, and is also not an effect size, since effect implies a relationship. Rather, the parameter being estimated could be called simply a single group summary. Note, however, that the classification of an index as an effect size and/or a treatment effect (or simply a single group summary) has no bearing on the computations.

In the meta-analysis itself we have simply a series of values and their variances, and the same mathematical formulas apply. In this volume we generally use the term effect size, but we use it in a generic sense, to include also treatment effects, single group summaries, or even a generic statistic.

2.3 CHOICE OF EFFECT SIZE

Borenstein (2009) and references within made an extensive review of the subject and presented all necessary theory. Three major considerations should drive the choice of an effect size index. The first is that the effect sizes from the different studies should be comparable to one another in the sense that they measure (at least approximately) the same thing. That is, the effect size should not depend on aspects of study design that may vary from study to study (such as sample size or whether covariates are used). The second is that estimates of the effect size should be computable from the information that is likely to be reported in published research reports. That is, it should not require the re-analysis of the raw data (unless these are known to be available). The third is that the effect size should have good technical properties. For example, its sampling distribution should be known so that variances and confidence intervals can be computed. Additionally, the effect size should be substantively interpretable. This means that researchers in the substantive area of the work represented in the synthesis should find the effect size meaningful. If the effect size is not inherently meaningful, it is usually possible to transform the effect size to another metric for presentation. For example, the analyses may be performed using the log risk ratio but then transformed to a risk ratio (or even to illustrative risks) for presentation.

In practice, the kind of data used in the primary studies will usually lead to a pool of two or three effect sizes that meet the criteria outlined above, which makes the process of selecting an effect size relatively straightforward. If the summary data reported by the primary study are based on means and standard deviations in two groups, the appropriate effect size will usually be either the raw difference in means, the standardized difference in means, or the response ratio. If the summary data are based on a binary outcome such as events and non-events in two groups the appropriate effect size will usually be the risk ratio, the odds ratio, or the risk difference. If the primary study reports a correlation between two variables, then the correlation coefficient itself may serve as the effect size.

2.4 PARAMETERS AND ESTIMATES

Borenstein (2009) and references within made an extensive review of the subject and presented all necessary theory. Throughout this volume we make the distinction between an underlying effect size parameter (denoted by the Greek letter θ) and the sample estimate of that parameter (denoted by Y). If a study had an infinitely large sample size then it would yield an effect size Y that was identical to the population parameter θ. In fact, though, sample sizes are finite and so the effect size estimate Y always differs from θ by some amount. The value of Y will vary from sample to sample, and the distribution of these values is the sampling distribution of Y. Statistical theory allows us to estimate the sampling distribution of effect size estimates, and hence their standard errors.

Effect sizes based on means

Raw (unstandardized) mean difference

Based on studies with independent groups

Based on studies with matched groups or pre-post designs

Standardized mean difference

Based on studies with independent groups

Based on studies with matched groups or pre-post designs

Response ratios

Based on studies with independent groups

Effect sizes based on binary data

Risk ratio (RR)

Based on studies with independent groups

Odds ratio (OR)

Based on studies with independent groups

Risk difference (RD)

Based on studies with independent groups

Effect sizes based on correlational data

Correlation

Based on studies with one group

2.5 A GENERAL FIXED EFFECTS PARAMETRIC APPROACH

2.5.1 A FIXED EFFECTS META-ANALYSIS MODEL

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. Suppose that there are *r* independent studies each comparing the treated group with the control group. There is a common outcome measure reported for each patient. The parameter representing the measure of treatment difference is denoted by θ. This may, for example, be the difference between treatment means for normally distributed data or the log-odds ratio for binary data. It is assumed here that θ equals 0 when the two treatments have equal effect. Denote by $\hat{\theta}_i$ an estimate of θ from the *i*th study. The general fixed effects model is given by:

$$\hat{\theta}_i = \theta + \varepsilon_i,$$

for $i = 1, \ldots, r$, where the ε_i are error terms and are realizations of normally distributed random variables with expected value 0 and variance denoted by ξ_i^2. It follows that

$$\hat{\theta}_i \sim N(\theta, \xi_i^2).$$

2.5.2 ESTIMATION AND HYPOTHESIS TESTING OF THE TREATMENT DIFFERENCE

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. Usually, the estimated variance of $\hat{\theta}_i$ var $(\hat{\theta}_i)$, is treated as if it were the true variance ξ_i^2 , that is, no allowance is made for error in the calculated term var$(\hat{\theta}_i)$ Let w_i be the estimated inverse variance of $\hat{\theta}_i$,that is, $w_i = 1/$ var $(\hat{\theta}_i)$. The distributional assumption that is made is that

$$\hat{\theta}_i \sim N(\theta, w_i^{-1}),$$

for $i = 1, \ldots, r$. Under the null hypothesis that the treatment difference in each study is equal to 0,

$$\hat{\theta}_i w_i \sim N(0, w_i),$$

for $i = 1, \ldots, r$, and, as the study estimates are independent,

$$\sum_{i=1}^{r} \hat{\theta}_i w_i \sim N\left(0, \sum_{i=1}^{r} w_i\right).$$

The global null hypothesis that the treatment difference in all studies is equal to 0 is tested by comparing the statistic

$$U = \frac{\left(\sum_{i=1}^{r} \hat{\theta}_i w_i\right)^2}{\sum_{i=1}^{r} w_i}$$

with the chi-squared distribution with one degree of freedom. Assuming that there is a common treatment difference in all studies,

$$\sum_{i=1}^{r} \hat{\theta}_i w_i \sim N\left(\theta \sum_{i=1}^{r} w_i, \sum_{i=1}^{r} w_i\right)$$

and the overall fixed effect θ can be estimated by ˆθ, where

$$\hat{\theta} = \frac{\sum_{i=1}^{r} \hat{\theta}_i w_i}{\sum_{i=1}^{r} w_i}.$$

If *w*i were the true inverse variance of $\hat{\theta}_i$, rather than being an estimate, then $\hat{\theta}$ would be the maximum likelihood estimate of θ. The standard error of $\hat{\theta}$ is given by

$$\text{se}(\hat{\theta}) = \sqrt{\frac{1}{\sum_{i=1}^{r} w_i}},$$

and an approximate 95% confidence interval (CI) for θ is given by

$$\hat{\theta} \pm 1.96\sqrt{\frac{1}{\sum_{i=1}^{r} w_i}}.$$

The calculations require an estimate of the treatment difference and its variance from each study. Usually a trial report will quote the standard error, and then w_i can be calculated as $1/\{\text{se}(\hat{\theta}_i)\}^2$. If using efficient score and Fisher's information

statistics, $\hat{\theta}_i = Z_i/V_i$. For this choice of $\hat{\theta}_i$ it follows that $w_i = V_i$. Also$\hat{\theta}_i w_i = Z_i$ and $\hat{\theta}_i{}^2 w_i = Z_i{}^2/V_i$. Thus

$$\hat{\theta} = \frac{\sum_{i=1}^{r} Z_i}{\sum_{i=1}^{r} V_i}$$

\

And

$$U = \frac{\left(\sum_{i=1}^{r} Z_i\right)^2}{\sum_{i=1}^{r} V_i}.$$

The fixed effects approach is sometimes referred to as an 'assumption-free' approach (see, for example, Early Breast Cancer Trialists' Collaborative Group, 1990) because it is argued that the fixed effects estimate does not rely on the assumption of a common treatment difference parameter across all studies. Suppose that the assumption of a common treatment difference in all studies is relaxed and that the distributional assumption for the individual study estimates becomes

$$\hat{\theta}_i \sim N(\theta_i, w_i^{-1}),$$

where θi is the treatment difference parameter in study *i*. The overall fixed effect estimate $\hat{}\theta$ can now be viewed as an estimate of

$$\frac{\sum_{i=1}^{r} \theta_i w_i}{\sum_{i=1}^{r} w_i},$$

the weighted mean of the study treatment difference parameters. Whilst this is an acceptable interpretation of $\hat{\theta}$, it would not appear to go far enough. Once variation between studies is conceded it would seem natural to investigate the amount of heterogeneity and to allow for it when making inferences about the difference between the two treatments.

2.5.3 TESTING FOR HETEROGENEITY ACROSS STUDIES

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. To test for heterogeneity in the treatment difference parameter across the studies,
a large-sample test is used. This is based on the statistic

$$Q = \sum_{i=1}^{r} w_i(\hat{\theta}_i - \hat{\theta})^2,$$

which is a weighted sum of squares of the deviations of individual study estimates from the overall estimate (Cochran, 1954).When treatment difference parameters are homogeneous, Q follows a chi-squared distribution with $r - 1$ degrees of freedom. An easier and equivalent formula for calculation is given by

$$Q = \sum_{i=1}^{r} \hat{\theta}_i^2 w_i - U.$$

When using efficient score and Fisher's information statistics, Q can be written as

$$Q = \sum_{i=1}^{r} V_i \left(\frac{Z_i}{V_i} - \frac{\sum_{i=1}^{r} Z_i}{\sum_{i=1}^{r} V_i} \right)^2 = \sum_{i=1}^{r} \left(\frac{Z_i^2}{V_i} \right) - \frac{\left(\sum_{i=1}^{r} Z_i \right)^2}{\sum_{i=1}^{r} V_i}.$$

The test statistics U and Q and the estimate $\hat{\theta}$ and its standard error can be obtained by performing a weighted least-squares regression, in which the observed responses (y) are the study estimates of treatment difference, $\hat{\theta}_i$, and there are no explanatory variables, only a constant term. The weights (w) are the values w_i.

2.6 FIXED EFFECTS MODELS FOR BINARY DATA

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. The observation y_{ij} is assumed to be a realization of a random variable Y_{ij}, which has a binomial distribution with parameter p_{ij} and enominator $n_{ij} = 1$. If p_{ij} represents the probability of success for patient j in trial i,

then y_{ij} = 1 if the patient response is a 'success' and 0 if the response is a 'failure'. The expected value of Y_{ij} is p_{ij} and the variance $p_{ij}(1 - p_{ij})$.

In order to model the dependence of p_{ij} on the explanatory variables $x1, x2, \ldots, xq$, a transformation which maps the unit interval (0, 1) onto the real line $(-\infty,\infty)$ is used. This transformation is known as the link function. The natural choice for estimating odds ratios is the logit link function, given by

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right).$$

The logit link function leads to the linear logistic model

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha + \eta_{ij},$$

where α is the intercept and η_{ij} is a linear combination of explanatory variables. This model is an example of a generalized linear model, details of which can be found in Section A.6 of the Appendix. An analogy with the general linear model can be seen with $\log\{p_{ij}/(1 - p_{ij})\}$ replacing μ_{ij}.

The model which will provide an overall fixed effects estimate of treatment difference, includes study and treatment as covariates. It is given by

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha + \beta_{0i} + \beta_1 x_{1ij}. \tag{5.4}$$

The parameter β_1 represents the common log-odds ratio of success on treatment relative to control.

2.6.1 ESTIMATION AND HYPOTHESIS TESTING

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. Parameter estimates are obtained using the method of maximum likelihood, The standard error for a single parameter or a linear combination of the parameters can be calculated from the observed or expected Fisher's information matrix. Confidence intervals are based on asymptotic normality. Models are compared by means of the likelihood ratio test statistic, that is, the change

in deviance (−2 times the log-likelihood) between two models, one of which contains the parameter(s) of interest while the other is identical except that it does not contain the parameter(s) of interest. This test statistic is compared with the chi-squared distribution. Any package which fits a linear logistic regression model can be utilized. To test the null hypothesis that the treatment difference in all studies is equal to 0, model (5.4) is compared with a model which only contains the study effects, namely

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha + \beta_{0i}. \tag{5.5}$$

Model (5.4) has $r + 1$ degrees of freedom associated with the model terms and model (5.5) has r. The likelihood ratio statistic, equal to the change in deviance between the two models, is compared with the chi-squared distribution with one degree of freedom.

2.6.2 TESTING HETEROGENEITY ACROSS STUDIED

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. In order to perform a test for heterogeneity of the treatment difference parameter across studies it is necessary to fit the model which includes the study by treatment interaction term. This is given by

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha + \beta_{0i} + \beta_{1i}x_{1ij}, \tag{5.6}$$

which has $2r$ degrees of freedom associated with the model terms. The test for heterogeneity is a test of the study by treatment interaction term and involves the comparison of models (5.4) and (5.6). The change in deviance between these two models is compared with the chi-squared distribution on $r - 1$ degrees of freedom, in the same way as the Q statistic.

2.7 METHODS OF FIXED EFFECTS

The inverse variance (IV) method computes a weighted average of log-odds ratios or absolute risk differences using the inverse of the within-study variance as the study weight. In keeping with methods implemented in standard meta-analytical software, the variance of the individual trial odds ratio was approximated by the method of Woolf while the variance of the risk difference was estimated using the Normal approximation. The DL method is similar, except that the study weight is equal to the inverse of the sum of the individual study's variance and the estimated among-study variance . This is therefore a random effects method (the only one we consider), and is identical to the IV method when the among-study variance is zero. The MH method combines odds ratios or risk differences, their pooled variance estimators being the unconditional product binomial. The Peto one-step method computes an approximation of the log-odds from the ratio of the efficient score to the Fisher information, both evaluated under the null hypothesis. These quantities are estimated, respectively, by the sum of the differences between the observed and expected numbers of events in the treatment arm and by the sum of the conditional hypergeometric variances. For the MH, inverse variance, DL and Peto methods, we have used the Wald z-test method for hypothesis tests and computation of 95 per cent confidence limits. The Wald test for an MH-combined odds ratio differs from the score test proposed by Mantel and Haenszel . The score test is mathematically, although not conceptually, identical to the Wald test for the Peto odds ratio. With rare outcomes, a trial will sometimes contain no events in one or both arms, which yields problems when computations involve ratios or reciprocals of numbers of events. For many methods, statistical packages routinely add 0.5 to each cell from such trials before metaanalysis to avoid divide-by-zero errors. When no event occurs in either arm of a trial such that a and c are both zero, the trial odds ratio is undefined, as the trial provides no information about either the likely direction or magnitude of the effect. We have excluded such trials from meta-analyses of odds ratios for all methods except the crude unstratified method. The trial's risk difference is defined in this situation (as zero), and thus these trials are included for analyses on the absolute risk scale even though they are excluded from analyses on the odds ratio scale

2.7.1 INVERSE VARIANCE METHOD

Inverse variance methods may be used to pool either binary or continuous data. In the general formula below, the effect size, denoted i , could be the log odds ratio, log relative risk, risk difference, difference in means or standardised mean difference from the ith trial(Egger et all., 2013). The effect sizes are combined to give a pooled estimate by calculating a weighted average of the treatment effects from the individual trials:

$$\theta_{\mathrm{IV}} = \frac{\sum w_i \theta_i}{\sum w_i}.$$

The weights are the reciprocals of the squared standard errors:

$$w_i = \frac{1}{\mathrm{SE}(\theta_i)^2}.$$

Thus larger studies, which have smaller standard errors, are given more weight than smaller studies, which have larger standard errors. This choice of weight minimises the variability of the pooled treatment effect IV. The standard error of IV is given by S

$$\mathrm{SE}(\theta_{\mathrm{IV}}) = \frac{1}{\sqrt{\sum w_i}}.$$

The heterogeneity statistic is given by

$$Q = \sum w_i (\theta_i - \theta_{\mathrm{IV}})^2.$$

The strength of this approach is its wide applicability. It can be used to combine any estimates that have standard errors available. Thus it can be used for estimates from many types of study, including standardized mortality ratios, diagnostic test indices, hazard ratios, and estimates from cross-over trials and cluster-randomized trials. It is also possible to use this method when crude 2*2 tables cannot be obtained for each study, but treatment effects and confidence intervals are available.

2.7.2 MANTEL-HAENSZEL METHOD

When data are sparse, both in terms of event rates being low and trials being small, the estimates of the standard errors of the treatment effects that are used in the inverse variance methods may be poor (Egger et all., 2013). Mantel–Haenszel methods use an alternative weighting scheme, and have been shown to be more robust when data are sparse, and may therefore be preferable to the inverse variance method. In other situations they give similar estimates to the inverse variance method. They are available only for binary outcomes. For each study, the effect size from each trial i is given weight w_i in the analysis. The overall estimate of the pooled effect, θ_{MH} is given by:

$$\theta_{\mathrm{MH}} = \frac{\sum w_i \theta_i}{\sum w_i}.$$

Unlike with inverse variance methods, relative effect measures are combined in their natural scale, although their standard errors (and confidence intervals) are still computed on the log scale. For combining odds ratios, each study's. OR is given weight

$$w_i = \frac{b_i c_i}{N_i},$$

and the logarithm of OR_{MH} has standard error given by

$$\mathrm{SE}[\ln(OR_{\mathrm{MH}})] = \sqrt{\frac{1}{2}\left(\frac{E}{R^2} + \frac{F+G}{R \times S} + \frac{H}{S^2}\right)},$$

Where

$$R = \sum \frac{a_i d_i}{N_i};\ S = \sum \frac{b_i c_i}{N_i};$$

$$E = \sum \frac{(a_i + d_i) a_i d_i}{N_i^2};\ F = \sum \frac{(a_i + d_i) b_i c_i}{N_i^2};$$

$$G = \sum \frac{(b_i + c_i) a_i d_i}{N_i^2};\ H = \sum \frac{(b_i + c_i) b_i c_i}{N_i^2}.$$

For combining risk ratios, each study's RR is given weight

$$w_i = \frac{c_i n_{1i}}{N_i},$$

and the logarithm of RR_{MH} has standard error given by

$$\mathrm{SE}[\ln(RR_{\mathrm{MH}})] = \sqrt{\frac{P}{R \times S}},$$

Where

$$P = \sum \frac{(n_{1i} n_{2i} (a_i + c_i) - a_i c_i N_i)}{N_i^2};\ R = \sum \frac{a_i n_{2i}}{N_i};\ S = \sum \frac{c_i n_{1i}}{N_i}.$$

For risk differences, each study's RD has the weight

$$w_i = \frac{n_{1i} n_{2i}}{N_i},$$

and RD_{MH} has standard error given by

$$\mathrm{SE}(RD_{\mathrm{MH}}) = \sqrt{\mathcal{J} / K^2},$$

where

$$\mathcal{J} = \sum \left(\frac{a_i b_i n_{2i}^3 + c_i d_i n_{1i}^3}{n_{1i} n_{2i} N_i^2} \right);\ K = \sum \left(\frac{n_{1i} n_{2i}}{N_i} \right).$$

However, the test of homogeneity is based upon the inverse variance weights and not the Mantel–Haenszel weights. The heterogeneity statistic is given by

$$Q = \sum w_i (\theta_i - \theta_{\mathrm{MH}})^2$$

where is the log odds ratio, log relative risk or risk difference.

2.7.3 PETO METHOD

An alternative to the Mantel–Haenszel method is a method due to Peto, (Egger et all., 2013),(sometimes attributed to Yusuf, or to Yusuf and Peto). The overall odds ratio is given by

$$OR_{\text{Peto}} = \exp\left(\frac{\sum w_i \ln(OR_i)}{\sum w_i}\right),$$

where the odds ratio OR_i is calculated using the approximate Peto method described in the individual trial section, and the weight w_i is equal to the hypergeometric variance of the event count in the intervention group, v_i . The logarithm of the odds ratio has standard error

$$\text{SE}[\ln(OR_{\text{Peto}})] = \frac{1}{\sqrt{\sum v_i}}.$$

The heterogeneity statistic is given by

$$Q = \sum v_i (\ln OR_i - \ln OR_{\text{Peto}})^2.$$

The approximation upon which Peto's method relies has shown to fail when treatment effects are very large, and when the sizes of the arms of the trials are seriously unbalanced. Severe imbalance, with, for example, four or more times as many participants in one group than the other, would rarely occur in randomised trials. In other circumstances, including when event rates are very low, the method performs well. Corrections for zero cell counts are not necessary for this method .

2.8 A GENERAL RANDOM EFFECTS PARAMETRIC APPROACH

2.8.1 A RANDOM EFFECT META-ANALYSIS MODEL

In a random effects model it is assumed that the treatment difference parameters in the r studies $(\theta 1, \ldots, \theta r)$ are a sample of independent observations from $N(\theta, \tau^2)$. The general random effects model is given by

$$\hat{\theta}_i = \theta + \nu_i + \varepsilon_i,$$

for $i = 1, \ldots, r$, where the ν_i are normally distributed random effects with mean 0 and variance τ^2. The terms ν_i and ε_i are assumed to be independently distributed. It follows that

$$\hat{\theta}_i \sim N(\theta, \xi_i^2 + \tau^2).$$

2.8.2 ESTIMATION AND HYPOTHESIS TESTING OF THE TREATMENT DIFFERENCE

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. Usually τ^2 is unknown and must be estimated from the data. Therefore, the distributional assumption that is made is that

$$\hat{\theta}_i \sim N(\theta, w_i^{-1} + \hat{\tau}^2),$$

where $\hat{\tau}^2$ is an estimate of τ^2. By setting

$$w_i^* = (w_i^{-1} + \hat{\tau}^2)^{-1},$$

it follows that

$$\hat{\theta}_i \sim N(\theta, (w_i^*)^{-1}).$$

Treating the term $(w^*_i)^{-1}$ as if it were the true variance of $\hat{\theta}_i$ provides the test statistic

$$U^* = \frac{\left(\sum_{i=1}^{r} \hat{\theta}_i w_i^*\right)^2}{\sum_{i=1}^{r} w_i^*},$$

which follows a chi-squared distribution with one degree of freedom under the null hypothesis of no treatment difference ($\theta = 0$). If $(w^*_i)^{-1}$ is the true variance of $\hat{\theta}_i$, then the ML estimate of θ is given by $\hat{\theta}^*$, where

$$\hat{\theta}^* = \frac{\sum_{i=1}^{r} \hat{\theta}_i w_i^*}{\sum_{i=1}^{r} w_i^*}.$$

Now $\hat{\theta}^*$ is asymptotically unbiased for θ, with variance approximately equal to $1/\sum_{i=1}^{r} w^*_i$. The standard error is given by

$$\text{se}(\hat{\theta}^*) = \sqrt{\frac{1}{\sum_{i=1}^{r} w_i^*}},$$

and an approximate 95% CI for θ is given by

$$\hat{\theta}^* \pm 1.96\sqrt{\frac{1}{\sum_{i=1}^{r} w_i^*}}.$$

If $\hat{\tau}^2$ is small then the modified weights w^*_i will be close to the original weights w_i. In this case the standard error and CI obtained from the random effects model will be similar to those from the fixed effects model. Also the overall estimate of treatment difference from both models will be similar. If $\hat{\tau}^2$ is large then the standard error and CI will be much larger for the random effects model. The random effects estimate of treatment difference will move closer towards the arithmetic mean of the individual study estimates. How much this estimate differs from the fixed effects estimate will depend on the extent to which the studies with the largest original weights w_i are associated with the extreme estimates of treatment difference.

2.8.3 ESTIMATION OF τ^2 USING THE METHOD OF MOMENTS

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. The approach to the estimation of τ^2 considered here is that based on the method of

moments. This estimate can be readily calculated without the need for a statistical software package. The following considerations provide the method of moments estimate for τ^2. Under the random effects model, the fixed effects estimate of θ,

$$\hat{\theta} = \frac{\sum_{i=1}^{r} \hat{\theta}_i w_i}{\sum_{i=1}^{r} w_i},$$

still has mean θ, but its variance is now given by

$$\text{var}(\hat{\theta}) = \frac{\sum_{i=1}^{r} w_i^2 \text{var}(\hat{\theta}_i)}{\left(\sum_{i=1}^{r} w_i\right)^2} = \frac{\sum_{i=1}^{r} w_i^2 (w_i^{-1} + \tau^2)}{\sum_{i=1}^{r} w_i^2}$$

$$= \frac{1}{\sum_{i=1}^{r} w_i} + \frac{\tau^2 \sum_{i=1}^{r} w_i^2}{\left(\sum_{i=1}^{r} w_i\right)^2}.$$

The statistic Q used for testing heterogeneity is

$$Q = \sum_{i=1}^{r} w_i (\hat{\theta}_i - \hat{\theta})^2 = \sum_{i=1}^{r} w_i (\hat{\theta}_i - \theta)^2 - \left(\sum_{i=1}^{r} w_i\right) (\hat{\theta} - \theta)^2,$$

so that the expected value of Q, E(Q), is given by

$$\text{E}(Q) = \sum_{i=1}^{r} w_i \text{var}(\hat{\theta}_i) - \left(\sum_{i=1}^{r} w_i\right) \text{var}(\hat{\theta})$$

$$= \sum_{i=1}^{r} w_i (w_i^{-1} + \tau^2) - \left(\sum_{i=1}^{r} w_i\right) \left\{ \frac{1}{\sum_{i=1}^{r} w_i} + \frac{\tau^2 \sum_{i=1}^{r} w_i^2}{\left(\sum_{i=1}^{r} w_i\right)^2} \right\}$$

$$= (r-1) + \tau^2 \left(\sum_{i=1}^{r} w_i - \frac{\sum_{i=1}^{r} w_i^2}{\sum_{i=1}^{r} w_i} \right).$$

This motivates use of the method of moments estimate $\hat{\tau}^2$ for τ^2, where

$$\hat{\tau}^2 = \frac{Q - (r-1)}{\sum_{i=1}^{r} w_i - \sum_{i=1}^{r} w_i^2 / \sum_{i=1}^{r} w_i},$$

as described by DerSimonian and Laird (1986). Because of the possibility of a negative method of moments estimate, in practice the estimate used is the maximum of the values 0 and $\hat{\tau}^2$. This means that when Q is smaller than its degrees of freedom the method of moments estimate will be set equal to 0. The test for heterogeneity, using Q, is a test of H_0: $\tau^2 = 0$. Should $\hat{\tau}^2 \leq 0$, a fixed effects analysis is more appropriate, because this happens when $Q < \mathrm{E}\,(Q; \tau 2 = 0) = r - 1$. It can be seen that setting τ^2=0 in the random effects model leads to the fixed effects model. If $\hat{\tau}^2 > 0$ the following approximate result may be used

$$\hat{\theta}_i \sim N(\theta, w_i^{-1} + \hat{\tau}^2) \equiv N(\theta, (w_i^*)^{-1}).$$

In a similar way, the test statistic $U*$ and the estimate

$\hat{\theta}^*$ and its standard error can be obtained by performing a weighted least-squares regression. The only difference is that for the random effects analysis the weights are the values w^*_i instead of w_i.

2.9 RANDOM EFFECTS MODELS FOR BINARY DATA

Borenstein (2009) and references within made an extensive review of the subject and presented all necessary theory. The fixed-effect model, discussed above, starts with the assumption that the true effect size is the same in all studies. However, in many systematic reviews this assumption is implausible. When we decide to incorporate a group of studies in a meta-analysis, we assume that the studies have enough in common that it makes sense to synthesize the information, but there is generally no reason to assume that they are identical in the sense that the true effect size is exactly the same in all the studies.

For example, suppose that we are working with studies that compare the proportion of patients developing a disease in two groups (vaccinated versus placebo). If the treatment works we would expect the effect size (say, the risk ratio) to be similar but not identical across studies. The effect size might be higher (or lower) when the

participants are older, or more educated, or healthier than others, or when a more intensive variant of an intervention is used, and so on. Because studies will differ in the mixes of participants and in the implementations of interventions, among other reasons, there may be different effect sizes underlying different studies.

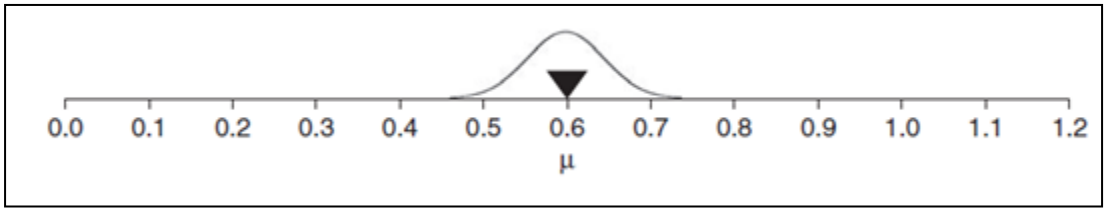

Figure 2.1: Random effects model- distribution of true effects.

Or, suppose that we are working with studies that assess the impact of an educational intervention. The magnitude of the impact might vary depending on the other resources available to the children, the class size, the age, and other factors, which are likely to vary from study to study. We might not have assessed these covariates in each study. Indeed, we might not even know what covariates actually are related to the size of the effect. Nevertheless, logic dictates that such factors do exist and will lead to variations in the magnitude of the effect.

One way to address this variation across studies is to perform a random-effects meta-analysis. In a random-effects meta-analysis we usually assume that the true effects are normally distributed. For example, in Figure 2.1 Borenstein (2009) , the mean of all true effect sizes is 0.60 but the individual effect sizes are distributed about this mean, as indicated by the normal curve. The width of the curve suggests that most of the true effects fall in the range of 0.50 to 0.70.

Suppose that our meta-analysis includes three studies drawn from the distribution of studies depicted by the normal curve, and that the true effects (denoted θ_1, θ_2, and θ_3) in these studies happen to be 0.50, 0.55 and 0.65 (see Figure 2.2 Borenstein (2009)).

If each study had an infinite sample size the sampling error would be zero and the observed effect for each study would be the same as the true effect for that study.

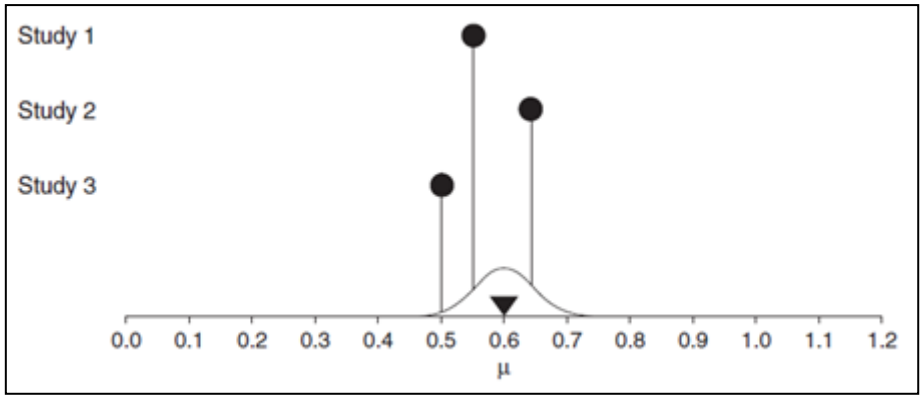

Figure 2.2: Random effects model- true effects.

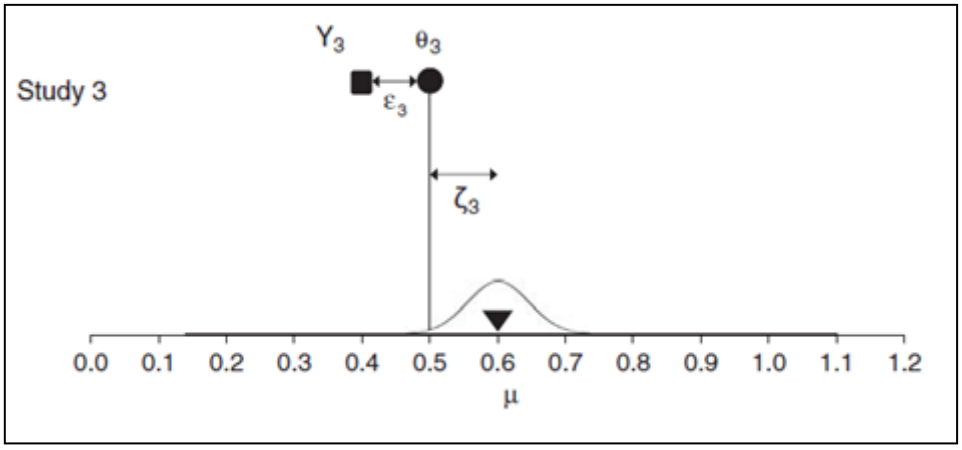

Figure 2.3: Random effects model- true and observed effect in one study.

If we were to plot the observed effects rather than the true effects, the observed effects would exactly coincide with the true effects. Of course, the sample size in any study is not infinite and therefore the sampling error is not zero. If the true effect size for a study is θ_i, then the observed effect for that study will be less than or greater than θ_i because of sampling error. For example, consider Study 3 in Figure 2.2. This study is the subject of Figure 2.3, where we consider the factors that control the observed effect. The true effect for Study 3 is 0.50 but the sampling error for this study is –0.10, and the observed effect for this study is 0.40.

This figure also highlights the fact that the distance between the overall mean and the observed effect in any given study consists of two distinct parts: true variation in effect sizes (ζ_i) and sampling error (e_i). In Study 3 the total distance from μ to Y_3 is -0.20. The distance from μ to θ_3 (0.60 to 0.50) reflects the fact that the true effect size actually varies from one study to the next, while the distance from θ_3 to Y_3 (0.5 to 0.4) is sampling error.

More generally, the observed effect Y_i for any study is given by the grand mean, the deviation of the study's true effect from the grand mean, and the deviation of the study's observed effect from the study's true effect. That is,

$$Y_i = \mu + \zeta_i + \varepsilon_i.$$

Therefore, to predict how far the observed effect Y_i is likely to fall from μ in any given study we need to consider both the variance of ζ_i and the variance of ε_i. The distance from μ (the triangle) to each θ_i (the circles) depends on the standard deviation of the distribution of the true effects across studies, called τ (tau) (or τ^2 for its variance). The same value of τ^2applies to all studies in the meta-analysis, and in Figure 2.4 Borenstein (2009) is represented by the normal curve at the bottom, which extends roughly from 0.50 to 0.70.

The distance from θ_i to Y_i depends on the sampling distribution of the sample effects about θ_i. This depends on the variance of the observed effect size from each study, V_{Yi} , and so will vary from one study to the next.

In Figure 2.4 Borenstein (2009) the curve for Study 1 is relatively wide while the curve for Study 2 is relatively narrow.

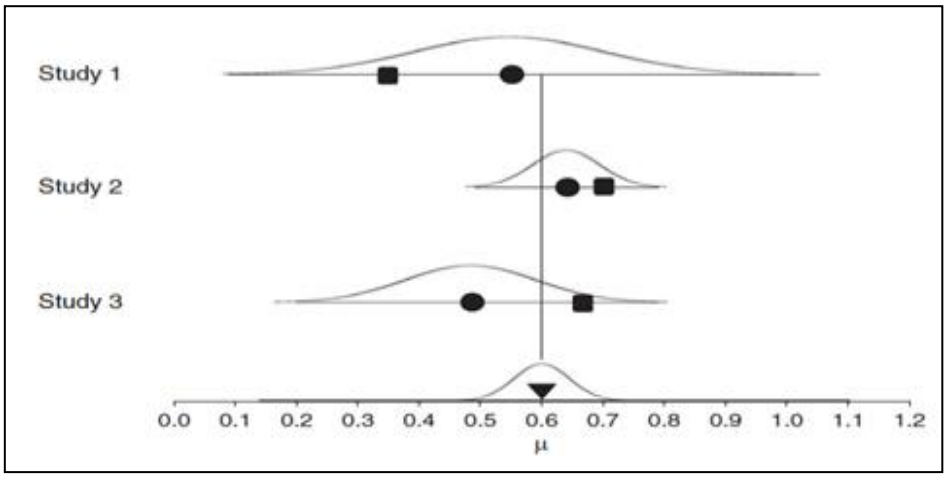

Figure 2.4: Random effects model- between study and within study variance.

2.9.1 A RANDOM EFFECTS META-ANALYSIS MODEL

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. The random effects meta-analysis model for the binary response in which the logit

link function is to be used is given by

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha + \beta_{0i} + \beta_1 x_{1ij} + \nu_{1i} x_{1ij},$$

and has been discussed by Turner *et al*. (2000). This model is an example of a generalized linear mixed model.

2.9.2 ESTIMATION AND HYPOTHESIS TESTING

Whitehead (2002) and references within made an extensive review of the subject and presented all necessary theory. The methodology and the software for fitting generalized linear mixed models has recently been and still is undergoing development. For a full maximum likelihood analysis based on the joint marginal distribution, numerical integration techniques are required for calculation of the log-likelihood, score equations and Fisher's information matrix. Approximate inference, which is available with the MLn program, involves the use of either marginal quasi-likelihood (MQL) or penalized quasi-likelihood (PQL), and either first-order or second-order Taylor expansion approximations for the logit link function. Approximate ML and REML estimates are found via the IGLS and RIGLS procedures. PQL produces improved estimates of variance components in mixed models, in general, whilst model convergence is more easily achieved with MQL. The second-order Taylor expansion provides greater accuracy than the first-order expansion. For further details about generalized linear mixed models, the reader is referred to Brown and Prescott (1999).

Wald tests can be used for inferences concerning the variance components. Wald tests can be used for inferences concerning the variance components. However, likelihood ratio tests based on the REML are preferable. Wald tests can be used for inferences concerning the fixed effect parameters. However, the calculated standard errors of the parameter estimates and the corresponding CIs are usually too narrow, because no allowance is made for the estimation of the variance components. Within MLn parametric bootstrapping may be used.

2.10 FIXED EFFECTS VS RANDOM EFFECTS

2.10.1 ESTIMATING THE SUMMARY EFFECT

Borenstein (2009) and references within made an extensive review of the subject and presented all necessary theory. Under the fixed-effect model we assume that the true effect size for all studies is identical, and the only reason the effect size varies between studies is sampling error (error in estimating the effect size). Therefore, when assigning weights to the different studies we can largely ignore the information in the smaller studies since we have better information about the same effect size in the larger studies.

By contrast, under the random-effects model the goal is not to estimate one true effect, but to estimate the mean of a distribution of effects. Since each study provides information about a different effect size, we want to be sure that all these effect sizes are represented in the summary estimate. This means that we cannot discount a small study by giving it a very small weight (the way we would in a fixed-effect analysis). The estimate provided by that study may be imprecise, but it is information about an effect that no other study has estimated. By the same logic we cannot give too much weight to a very large study (the way we might in a fixed-effect analysis). Our goal is to estimate the mean effect in a range of studies, and we do not want that overall estimate to be overly influenced by any one of them.

2.10.2 CONFIDENCE INTERVAL

Borenstein (2009) and references within made an extensive review of the subject and presented all necessary theory. Under the fixed-effect model the only source of uncertainty is the within-study (sampling or estimation) error. Under the random-effects model there is this same source of uncertainty plus an additional source (between-studies variance). It follows that the variance, standard error, and confidence interval for the summary effect will always be larger (or wider) under the random-effects model than under the fixed-effect model (unless τ^2 is zero, in which case the two models are the same).

2.10.3 THE NULL HYPOTHESIS

Borenstein (2009) and references within made an extensive review of the subject and presented all necessary theory. Often, after computing a summary effect, researchers perform a test of the null hypothesis. Under the fixed-effect model the null hypothesis being tested is that there is zero effect in every study. Under the random-effects model the null hypothesis being tested is that the mean effect is zero. Although some may treat these hypotheses as interchangeable, they are in fact different, and it is imperative to choose the test that is appropriate to the inference a researcher wishes to make.

2.10.4 WHICH MODEL SHALL WE USE?

Borenstein (2009) and references within made an extensive review of the subject and presented all necessary theory:

Fixed effect

It makes sense to use the fixed-effect model if two conditions are met. First, we believe that all the studies included in the analysis are functionally identical. Second, our goal is to compute the common effect size for the identified population, and not to generalize to other populations. For example, suppose that a pharmaceutical company will use a thousand patients to compare a drug versus placebo. Because the staff can work with only 100 patients at a time, the company will run a series of ten trials with 100 patients in each. The studies are identical in the sense that any variables which can have an impact on the outcome are the same across the ten studies. Specifically, the studies draw patients from a common pool, using the same researchers, dose, measure, and so on (we assume that there is no concern about practice effects for the researchers, nor for the different starting times of the various cohorts). All the studies are expected to share a common effect and so the first condition is met. The goal of the analysis is to see if the drug works in the population from which the patients were drawn (and not to extrapolate to other populations), and so the second condition is met, as well.

In this example the fixed-effect model is a plausible fit for the data and meets the goal of the researchers. It should be clear, however, that this situation is relatively rare. The vast majority of cases will more closely resemble those discussed immediately below.

Random effects

By contrast, when the researcher is accumulating data from a series of studies that had been performed by researchers operating independently, it would be unlikely that all the studies were functionally equivalent. Typically, the subjects or interventions in these studies would have differed in ways that would have impacted on the results, and therefore we should not assume a common effect size. Therefore, in these cases the random-effects model is more easily justified than the fixed-effect model. Additionally, the goal of this analysis is usually to generalize to a range of scenarios. Therefore, if one did make the argument that all the studies used an identical, narrowly defined population, then it would not be possible to extrapolate from this population to others, and the utility of the analysis would be severely limited.

A caveat

There is one caveat to the above. If the number of studies is very small, then the estimate of the between-studies variance will have poor precision. While the random-effects model is still the appropriate model, we lack the information needed to apply it correctly. In this case the reviewer may choose among several options, each of them problematic.

One option is to report the separate effects and not report a summary effect. The hope is that the reader will understand that we cannot draw conclusions about the effect size and its confidence interval. The problem is that some readers will revert to vote counting and possibly reach an erroneous conclusion. Another option is to perform a fixed-effect analysis. This approach would yield a descriptive analysis of the included studies, but would not allow us to make inferences about a wider population. The problem with this approach is that (a) we do want to make inferences about a wider population and (b) readers will make these inferences even if they are not warranted. A third option is to take a Bayesian approach, where the estimate is based on data from outside of the current set of studies. This is probably the best option, but the problem is that relatively few researchers have expertise in Bayesian meta-analysis.

The test of the null hypothesis between studies variance is zero, is based on the amount of between-studies variance observed, relative to the amount we would expect if the studies actually shared a common effect size. Some have adopted the practice of starting with a fixed-effect model and then switching to a random-effects

model if the test of homogeneity is statistically significant. This practice should be strongly discouraged because the decision to use the random-effects model should be based on our understanding of whether or not all studies share a common effect size, and not on the outcome of a statistical test (especially since the test for heterogeneity often suffers from low power). If the study effect sizes are seen as having been sampled from a distribution of effect sizes, then the random-effects model, which reflects this idea, is the logical one to use. If the between-studies variance is substantial (and statistically significant) then the fixed-effect model is inappropriate. However, even if the between-studies variance does not meet the criterion for statistical significance (which may be due simply to low power) we should still take account of this variance when assigning weights. If τ^2 turns out to be zero, then the random-effects analysis reduces to the fixed-effect analysis, and so there is no cost to using this model. On the other hand, if one has elected to use the fixed-effect model a priori but the test of homogeneity is statistically significant, then it would be important to revisit the assumptions that led to the selection of a fixed-effect model.

The discussion of differences between the fixed-model and the random-effects model focused largely on the computation of a summary effect and the confidence intervals for the summary effect. We did not address the implications of the dispersion itself. Under the fixed-effect model we assume that all dispersion in observed effects is due to sampling error, but under the random-effects model we allow that some of that dispersion reflects real differences in effect size across studies. In the chapters that follow we discuss methods to quantify that dispersion and to consider its substantive implications.

Although a fixed-effect meta-analysis is defined as assuming that every study has a common true effect size, some have argued that the fixed effect method is valid without making this assumption. The point estimate of the effect in a fixed-effect meta-analysis is simply a weighted average and does not strictly require the assumption that all studies estimate the same thing. For simplicity and clarity we adopt a definition of a fixed-effect meta-analysis that does assume homogeneity of effect.

2.11 STUDY SELECTION

For the study selection of studies Anne Whitehead in the book Meta-Analysis of controlled clinical trials, in an extended review of the subject presented all necessary theory. The selection criteria for studies in the meta-analysis should be specified. If there is more than one hypothesis to be tested it may be necessary to define separate selection criteria for each one. In addition, for each hypothesis of interest, it may be desirable to create two groups of studies. The first group would consist of the primary studies on which the formal meta-analysis would be undertaken. The second group would consist of additional studies whose results may be included in a sensitivity analysis, or in a graphical presentation of individual study results.

Such studies may involve different patient populations or treatment comparisons from the primary studies, or may have less appropriate designs. However, their results may still be informative. Careful thought needs to be given to the selection criteria for the primary studies.

If they are very strict, the results of the meta-analysis may only be applicable to a small subset of the patient population or to a very specific treatment regimen, whereas if they are too liberal, it may not be possible to combine the individual trial results in an informative way. Typically, the selection criteria will define the treatment of interest and the relevant subject population. This should follow logically from the statement of the objectives of the meta-analysis. In addition, they may relate to the type of study design used. The assessment of the methodological quality of a trial may also be used to determine its eligibility for inclusion in the group of primary studies. The most important aspect of this assessment concerns the avoidance of bias in the estimation of the treatment difference of interest. Therefore, design issues, such as the method of randomizing subjects to treatment group, blinding, method of assessing patient outcome, follow-up of patients, and handling of protocol deviations and patient withdrawals from the trial, are likely to feature prominently. It may be appropriate to categorize studies according to how well they adhere to important methodological standards. In the report of a meta-analysis it will be necessary to include a list of studies which were excluded as well as a list of studies which were included. The reason for exclusion should be provided for each excluded study. It may be advantageous to have more than one assessor decide independently which

studies to include or exclude, together with a well-defined checklist and a procedure which will be followed when they disagree. In some cases, new information may surface during the reading of the study reports which indicate a need to modify the study selection criteria.

2.12 THE PACKAGE "meta" and "metafor" IN R.

The meta package provides functions for conducting meta-analyses in R. In the r project site where all functions of every package in r software are described we found that the package includes functions for fitting the meta-analytic fixed- and random-effects models and allows for the inclusion of moderators variables (study-level covariates) in these models. Meta-regression analyses with continuous and categorical moderators can be conducted in this way. Functions for the Mantel-Haenszel and Peto's one-step method for meta-analyses of 2 × 2 table data are also available. Finally, the package provides various plot functions (for example, for forest, funnel, and radial plots) and functions for assessing the model fit, for obtaining case diagnostics, and for tests of publication bias.

Details R package meta (Schwarzer, 2007) provides the following meta-analysis methods:

• Fixed effect and random effects meta-analysis (functions metabin, metacont, metacor, metagen, metainc, metaprop, and metarate)

• Several plots (forest, funnel, Galbraith / radial, labbe, baujat, bubble)

• Statistical tests (metabias) and trim-and-fill method (trimfill) to evaluate bias in metaanalysis

• Import data from 'RevMan 5' (read.rm5; see also metacr)

• Prediction interval, Hartung-Knapp and Paule-Mandel method for random effects model (arguments in meta-analysis functions)

• Cumulative meta-analysis (metacum) and leave-one-out meta-analysis (metainf)

• Meta-regression (metareg; if R package metafor is installed)

• Generalised linear mixed models (metabin, metainc, metaprop, and metarate; if R packages metafor, lme4, numDeriv, and BiasedUrn are installed)

A comprehensive collection of functions for conducting meta-analyses in R. The package includes functions to calculate various effect sizes or outcome measures, fit fixed-, random- , and mixed-effects models to such data, carry out moderator and meta-regression analyses, and create various types of meta-analytical plots (e.g.,

forest, funnel, radial, L'Abbe, Baujat, GOSH plots). For meta-analyses of binomial and person-time data, the package also provides functions that implement specialized methods, including the MantelHaenszel method, Peto's method, and a variety of suitable generalized linear (mixedeffects) models (i.e., mixed-effects logistic and Poisson regression models). Finally, the package provides functionality for fitting meta-analytic multivariate/multilevel models that account for non-independent sampling errors and/or true effects (e.g., due to the inclusion of multiple treatment studies, multiple endpoints, or other forms of clustering). Network metaanalyses and meta-analyses accounting for known correlation structures (e.g., due to phylogenetic relatedness) can also be conducted.

2.13 FOREST PLOT

A forest plot, also known as a blobbogram, is a graphical display of estimated results from a number of scientific studies addressing the same question, along with the overall results. It was developed for use in medical research as a means of graphically representing a meta-analysis of the results of randomized controlled trials. In the last twenty years, similar meta-analytical techniques have been applied in observational studies (e.g. environmental epidemiology) and forest plots are often used in presenting the results of such studies also.

Although forest plots can take several forms, they are commonly presented with two columns. The left-hand column lists the names of the studies (frequently randomized controlled trials or epidemiological studies), commonly in chronological order from the top downwards. The right-hand column is a plot of the measure of effect (e.g. an odds ratio) for each of these studies (often represented by a square) incorporating confidence intervals represented by horizontal lines. The graph may be plotted on a natural logarithmic scale when using odds ratios or other ratio-based effect measures, so that the confidence intervals are symmetrical about the means from each study and to ensure undue emphasis is not given to odds ratios greater than 1 when compared to those less than 1. The area of each square is proportional to the study's weight in the meta-analysis. The overall meta-analysed measure of effect is often represented on the plot as a dashed vertical line. This meta-analysed measure of

effect is commonly plotted as a diamond, the lateral points of which indicate confidence intervals for this estimate.

S Lewis (2001) in an extended review of the subject presented all necessary theory based on the following ideas. In a typical forest plot, the results of component studies are shown as squares centred on the point estimate of the result of each study. A horizontal line runs through the square to show its confidence interval—usually, but not always, a 95% confidence interval. The overall estimate from the meta-analysis and its confidence interval are put at the bottom, represented as a diamond. The centre of the diamond represents the pooled point estimate, and its horizontal tips represent the confidence interval. Significance is achieved at the set level if the diamond is clear of the line of no effect.

The plot allows readers to see the information from the individual studies that went into the meta-analysis at a glance. It provides a simple visual representation of the amount of variation between the results of the studies, as well as an estimate of the overall result of all the studies together. Forest plots increasingly feature in medical journals, and the growth of the Cochrane Collaboration has seen the publication of thousands in recent years.

The origin of forest plots goes back at least to the 1970s. Freiman et al displayed the results of several studies with horizontal lines showing the confidence interval for each study and a mark to show the point estimate. This study was not a meta-analysis, and the results of the individual studies were therefore not combined into an overall result.[2] In 1982, Lewis and Ellis produced a similar plot but this time for a meta-analysis, and they put the overall effect on the bottom of the plot. However, smaller studies, with less precise estimates of effect, had larger confidence intervals and, perversely, were the most noticeable on the plots

In S. Gopalakrishnan and P. Ganeshkumar (2013) a forest plot is described at the following structure: meta-analysis graphs can principally be divided into six columns [Figure 2.5]. Individual study results are displayed in rows. The first column (“study”) lists the individual study IDs included in the meta-analysis; usually the first author and year are displayed. The second column relates to the intervention groups and the third column to the control groups. The fourth column visually displays the study results. The line in the middle is called “the line of no effect.” The weight (in %) in

the fifth column indicates the weighting or influence of the study on the overall results of the meta-analysis of all included studies. The higher the percentage weight, the bigger the box, the more influence the study has on the overall results. The sixth column gives the numerical results for each study (e.g., odds ratio or relative risk and 95% confidence interval), which are identical to the graphical display in the fourth column. The diamond in the last row of the graph illustrates the overall result of the meta-analysis.

Thus, the horizontal lines represent individual studies. Length of line is the confidence interval (usually 95%), squares on the line represent effect size (risk ratio) for the study, with area of the square being the study size (proportional to weight given) and position as point estimate (relative risk) of the study

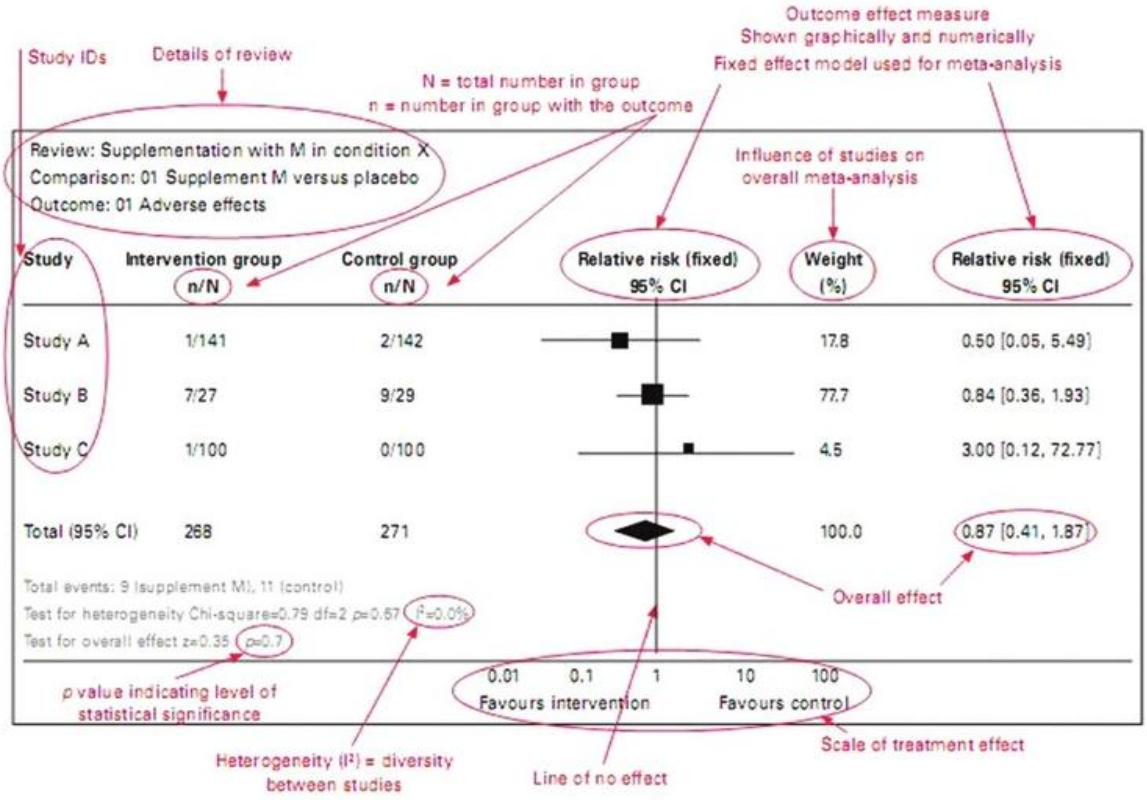

Figure 2.5: Interpretation of meta-analysis

An example is presented of how to read a forest plot that can be found at the centre for evidence-based intervention of the University of Oxford:

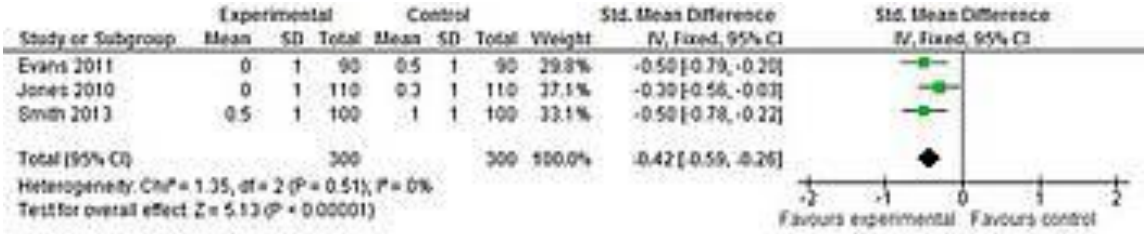

Figure 2.6: Forest plot of the example.

Interpreting the forest plot involves two steps:

Determine the effect size and

Assess the level of difference (or heterogeneity) among the different trials that are included in the meta-analysis

In the example, all of the lines fall on the left-hand side of the graph (Figure 2.6), which tells us that, in each of the trials, the participants who received the intervention showed or reported bigger changes than the participants who received the control condition (the control condition may have been another intervention or no intervention at all). The black diamond sits about half way between 0 and -1, which means that the average effect size of the three trials is about -0.5. For a more precise idea of the average effect size of the three trials, the actual number is reported in the table in boldface type, under the 'Std. Mean Difference' column. In this case, the actual average effect size is -0.42. According to a common interpretation of effect sizes, this would suggest that the intervention being tested in these three studies had a small to medium effect size – in other words, 'it worked' and had a moderate effect. In addition to the effect size, it is also important to consider the level of heterogeneity in a meta-analysis, which is captured in the I^2 statistic (which can be found at the bottom of the table in the example forest plot).

Systematic reviews and meta-analyses aim to capture the overall effects of an intervention or treatment when it has been tested in multiple trials. Ideally, if multiple trials are testing the same intervention, the effects of the intervention should be consistent across all of the studies. Unfortunately, this is rarely the case, because many things can affect the results of a trial, such as researcher bias, problems with data collection, or any number of other things.

So a systematic review and meta-analysis are designed to ask the question: If

these studies are all testing the same intervention, why don't they get the same results? Are the differences caused by chance, or is there something else involved? If it is the former, then we can have confidence in the results of the meta-analysis. If the differences are not the result of chance, then we need to be cautious in interpreting the results of the meta-analysis. Fortunately, it is easy to tell if heterogeneity is due to chance (or not) by interpreting the I^2 statistic. The I^2 statistic can be found at the bottom of the table in a forest plot. An I^2 statistic of more than 50% is considered high. In our example forest plot, $I^2 = 0\%$, so we can have confidence that the effects of the intervention being tested – which have a moderate effect size (-0.42) – are accurate and can be trusted. If the I^2 statistic were more than 50%, we would be less sure that the intervention can consistently have a moderate effect, and we might want to read the rest of the study to see if the authors report on why the effects are so different across studies. This can help you to determine, for example, with whom the intervention worked (e.g. who were the participants?) and to find out other details that might help you make a decision about whether the intervention has been tested with people or in places that are similar to your own population, clients or context.

2.14 FUNNEL PLOT

A funnel plot is a graph designed to check for the existence of publication bias; funnel plots are commonly used in systematic reviews and meta-analyses. In the absence of publication bias, it assumes that studies with high precision will be plotted near the average, and studies with low precision will be spread evenly on both sides of the average, creating a roughly funnel-shaped distribution.

Matthias Egger (1997) and references within made an extensive review of the subject and presented all necessary theory at the paper *"Bias in meta-analysis detected by a simple, graphical test"*. The authors mention that systematic reviews of the best available evidence regarding the benefits and risks of medical interventions can inform decision making in clinical practice and public health. Such reviews are, whenever possible, based on meta-analysis: "a statistical analysis which combines or integrates the results of several independent clinical trials considered by the analyst to be 'combinable'. However, the findings of some meta-analyses have later been contradicted by large randomized controlled trials. Such discrepancies have brought discredit on a technique that has been controversial since the outset. The appearance

of misleading meta-analysis is not surprising considering the existence of publication bias and the many other biases that may be introduced in the process of locating, selecting, and combining studies.

Funnel plots, plots of the trials' effect estimates against sample size, may be useful to assess the validity of meta-analyses. The funnel plot is based on the fact that precision in estimating the underlying treatment effect will increase as the sample size of component studies increases. Results from small studies will scatter widely at the bottom of the graph, with the spread narrowing among larger studies. In the absence of bias the plot will resemble a symmetrical inverted funnel. Conversely, if there is bias, funnel plots will often be skewed and asymmetrical.

The value of the funnel plot has not been systematically examined, and symmetry (or asymmetry) has generally been defined informally, through visual examination. Unsurprisingly, funnel plots have been interpreted differently by different observers. We measured funnel plot asymmetry numerically and examined the extent to which such asymmetry predicts discordance of results when meta-analyses are compared to single large trials of the same issue. We used the same method to assess the prevalence of funnel plot asymmetry, and thus of possible bias, among meta-analyses published in leading general medicine journals and meta-analyses disseminated electronically by the Cochrane Collaboration.

We present an example of how to read a funnel plot:

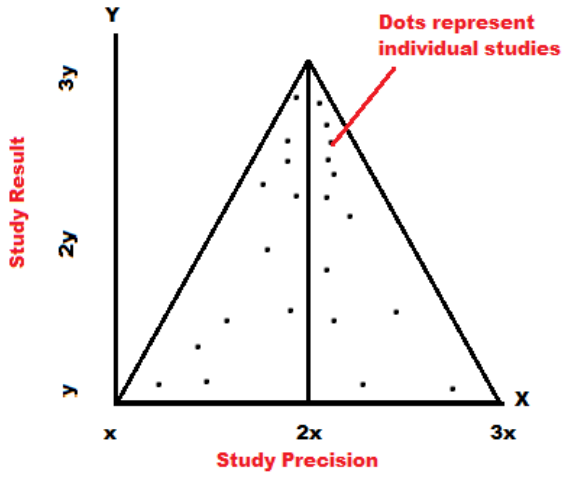

Figure 2.7: Structure of a funnel plot.

A funnel plot is a scatter plot of individual studies, their precision and results.
Funnel plots have the following characteristics:

Each dot represents a single study.
The y-axis is usually the standard error of the effect estimate. Larger studies with higher power are placed towards the top. Lower powered studies are placed towards the bottom. However, other measures could also be plotted (e.g. the reciprocal of the standard error, the reciprocal of the sample size, or variance of the estimated effect).
The x-axis shows the result for the study, sometimes expressed as an odds ratio.
The plot should ideally resemble a pyramid or inverted funnel, with scatter due to sampling variation. The shape is expected because the studies have a wide range of standard errors. If the standard errors were the same size, the studies would all fall on a horizontal line.
Funnel plots can be used as a check for bias in meta-analysis results. Asymmetry is commonly equated with publication bias and other kinds of reporting bias. However, funnel plots are not a good way to investigate publication bias (Sedgwick). There can be a number of reasons for asymmetrical funnel plots (also called *small study effects*). Sterne et. al (2011) list a slew of reasons, which include, but aren't limited to:
Poor methodological design, including fraud or inadequate analysis.
Reporting bias, including delayed publication and location bias, selective outcome reporting and selective analysis reporting. Can also include language bias (i.e. only including those studies written in your native language).
Chance: 95% of studies will usually fall within the triangular region if there are no biases or heterogeneity present in the studies. One possibility to skew the shape is that the errant 5% might all fall in one particular area by chance alone. The "95%" rule is actually a probability, meaning that chance alone could cause a higher or lower percentage than 95%, causing an asymmetrical shape that's actually not an indication of any bias at all. This is especially true if only a small number of studies are included in the meta analysis.
Study Heterogeneity. If heterogeneity results in a correlation between study size and intervention effects, this will result in an asymmetrical funnel (Terrin et. al)
The decision about whether a funnel plot is symmetric or not shouldn't be based only on visual cues. Tests for asymmetry are available (one such test is Egger's test), but

they should be interpreted with caution. They may not have statistical validity, typically have low power, and they may be challenging to interpret.

For the data analysis and the graphs R statistical software was used and the packages meta and metaphor as described in Chapter 2. In the introduction the problem was described as a problem of meta-analysis in studies where oral health problems are compared between people with severe mental illness and a control group.

CHAPTER 3: STATISTICAL METHODS AND ANALYSIS

In this chapter the search strategy for the paper's collection, the statistical techniques used and the analysis of the data are presented. As one of the aims of this thesis is to provide a comparison between three fixed effects methods (Mantel-Haenszel, Peto and Inverse variance), the results will be calculated using all these methods and a comparison between them will be provided.

3.1 SEARCH STRATEGY

The authors searched Medline, PsycINFO and Embase for the period January 1988 until March 2010 using the following text, MeSH or Emtree terms as appropriate: mental illness, mental disorders, dementia, psychosis, psychotic disorders, depression, depressive disorders, bipolar disorder, mood disorder, schizophrenia, oral health, dentistry and dental care. They searched for further publications by scrutinizing the reference lists of initial studies identified and other relevant review papers. For inclusion in the meta-analysis, studies had to have suitable controls. Where these were not included the authors looked for controls from a survey of a similar community and age group, conducted within 10 years of the index study. This is because oral health varies between populations, by both age and over time: for example, oral health has improved considerably over the past 20 years in most high-income countries. It was also ensured that the comparison data came from areas with similar levels of fluoride in the water supply.

3.2 STATISTICAL ANALYSIS

Package R software and libraries "meta" and "metafor" were used for the analysis. We calculated odds ratios, risk ratios and mean differences for edentulousness, given that the studies included had a crosssectional design. The authors (Kisely et all., 2011) calculated the mean differences for continuous data as studies used the same scale for each outcome (DMFT, DMFS). We assessed heterogeneity using the I^2 statistic. This provides an estimate of the percentage of

variability due to heterogeneity rather than chance alone. An I^2 estimate of 50% or greater indicates possible heterogeneity, and scores of 75–100% indicate considerable heterogeneity. The I^2 statistic is calculated using the chi-squared statistic (Q) and its degrees of freedom.

We used a fixed effects model with the Mandel Haenzel, Inverse Variance and Peto method for the calculation of and a random effects model for the calculation of odds ratios, risk ratios and mean differences for edentulousness, since we found significant heterogeneity in the majority of our analyses. In the other measurement(DMFS, Decayed teeth, missing teeth, Decayed Surfaces, DMFT) we used random effects models. For all measurements we performed a forest plot analysis for visual inspection of heterogeneity and a funnel plot analysis for graphical presentation of publication bias.

The authors found over 550 citations of interest in the initial electronic searches, and the final number of papers used is depicted into Figure2.1. Ten studies were from Europe; four were from India, three from Israel, two from Australia and one each from South Africa, Hong Kong and the USA. The most common diagnosis was psychosis, usually schizophrenia. Other diagnose (in descending order of frequency) included dementia, bipolar affective disorder, mood disorder, anxiety and personality disorder. Only seven studies used ICD or DSM diagnostic criteria. Ages ranged from 15 to 96 years, (Kisely et all., 2011).

3.3 INCLUSION AND EXCLUSION CRITERIA OF STUDIES

Studies were included with a focus on severe mental illness, defined as a primary diagnosis of dementia, schizophrenia, bipolar affective disorder or other affective disorder. Studies were included using clinical diagnoses or diagnostic criteria. Studies were excluded of eating disorder and of post-traumatic stress disorder in veterans, as these are very different patient groups. Studies were also excluded of people with primary alcohol or substance use disorders and people with intellectual disability for the same reason. Finally, the focus was on edentulousness as the end-stage of the two main dental diseases. Therefore there were also excluded studies of less severe dental outcomes such as poor oral hygiene. As a result, the finaly list of included studies was the following:

For the subgroup of edentulousness:

Adam,2006

Burchell,2006

Chalmers, 1998

Hede,1992

Hede,1995(35-49 years)

Hede,1995(65-78 years)

Lewis,2001

Mirza,2001

Tang,2004

Viglid,1993

For the measurement of decayed surfaces:

Hede, 1995 (35–49 years)

Hede, 1995 (65–78 years)

Stoefe; 1990

For measurement of DMFS:

Hede, 1995 (35–49 years)

Hede, 1995 (65–78 years)

Stiefel, 1990

For the measurement of decayed teeth:

Ramon, 2003 (18–34 years)

Velasco, 1997

For the measurement of Missing teeth:

Ramon, 2003 (18–34 years)

Stiefel, 1990

Velesco, 1997

For the measurement of DMFT:

Kumar, 2006

Ramon, 2003 (18–34 years)

Rekha, 2002

Velasco, 1997

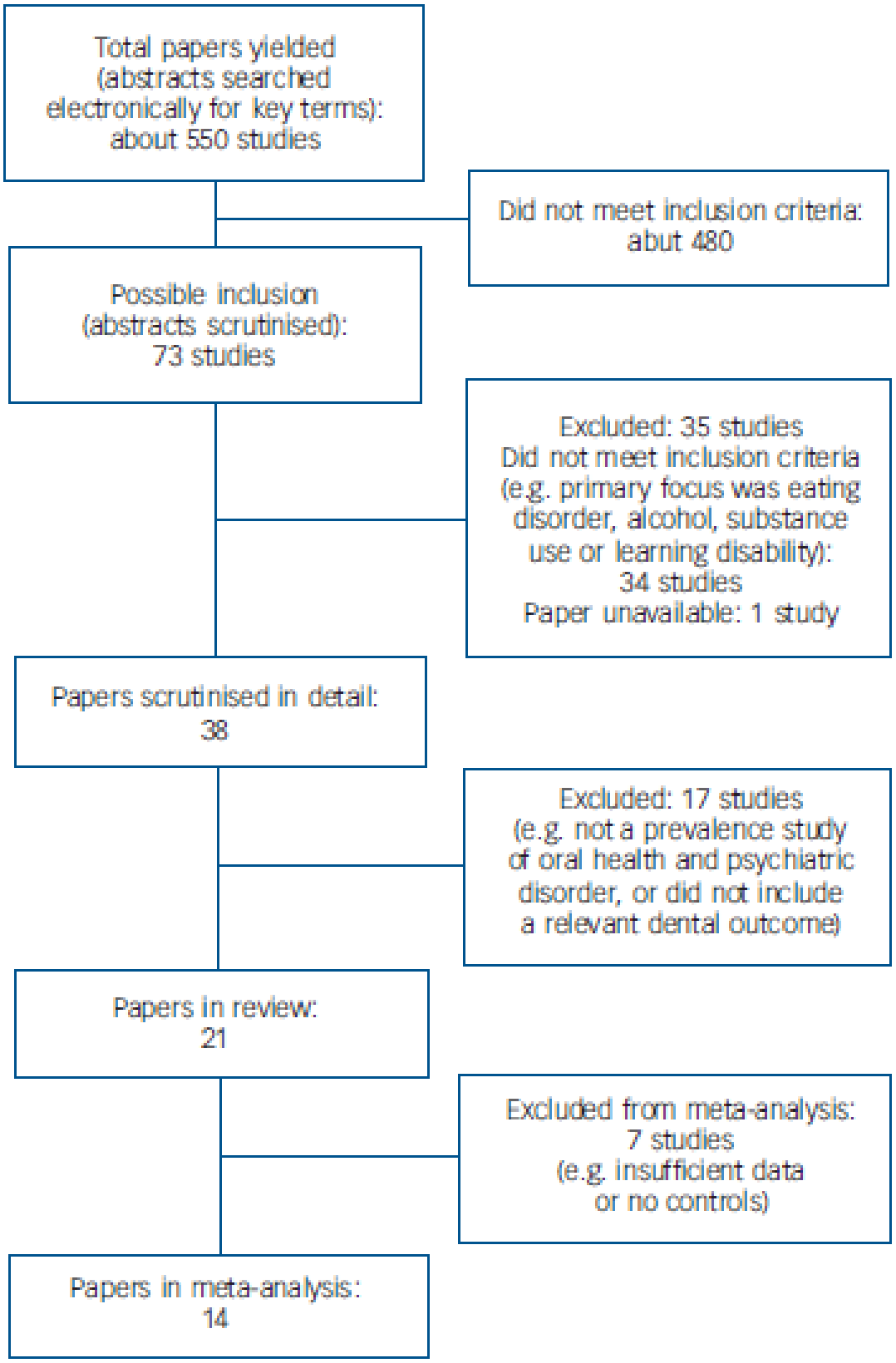

Figure 3.1: Papers yielded by search strategy in systematic review.

3.4 META-ANALYSIS RESULTS

To import the data in the R software we used the following commands:

data_psych <- data.frame(total=c(135,220,138,84,109,83,326,29,91,407),
events=c(89,16,14,23,2,18,205,2,6,256))

data_control<- data.frame(total=c(219,2667,3630,261,5759,8592,188,302,375,455),
events=c(109,213,200,8,58,1352,94,3,0,43))

study <- c('Adam,2006','Burchell,2006','Chalmers,1998','Hede,1992','Hede,1995(35-49years)','Hede,1995(65-78
years)','Lewis,2001','Mirza,2001','Tang,2004','Viglid,1993')

data_sbg1<-
data.frame(row.names=study,Psychiatric=data_psych,Control=data_control)

data_sbg1

	study	Psychiatric.total	Psychiatric.events	Control.total	Control.events
1	Adam,2006	135	89	219	109
2	Burchell,2006	220	16	2667	213
3	Chalmers,1998	138	14	3630	200
4	Hede,1992	84	23	261	8
5	Hede,1995(35-49 years)	109	2	5759	58
6	Hede,1995(65-78 years)	83	18	8592	1352
7	Lewis,2001	326	205	188	94
8	Mirza,2001	29	2	302	3
9	Tang,2004	91	6	375	0
10	Viglid,1993	407	256	455	43

Table 3.1: Data frame containing the data.

In order to do the meta-analysis in Odds Ratio we use the following command:

library(meta)

mOR<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_sbg1$Control.events,data_sbg1$Control.total,sm="OR" , method ="MH" , MH.exact = TRUE)

library(metafor)

forest(mOR,studlab=study)

```
> m
         OR                95%-CI %W(fixed) %W(random)
1    1.9525 [ 1.2529;     3.0428]      18.1        11.6
2    0.9036 [ 0.5331;     1.5317]      19.2        11.5
3    1.9363 [ 1.0943;     3.4263]       8.4        11.3
4   11.9242 [ 5.0880;    27.9453]       1.8        10.5
5    1.8373 [ 0.4430;     7.6201]       1.4         8.5
6    1.4829 [ 0.8770;     2.5075]      12.9        11.5
7    1.6942 [ 1.1777;     2.4372]      28.3        11.8
8    7.3827 [ 1.1819;    46.1166]       0.3         7.1
9   57.0936 [ 3.1857; 1023.2145]       0.0         4.4
10 16.2440 [11.1870;    23.5869]       9.6        11.8

Number of studies combined: k = 10

                           OR           95%-CI     z  p-value
Fixed effect model   3.2478 [2.7798; 3.7945] 14.84 < 0.0001
Random effects model 3.3510 [1.5703; 7.1511]  3.13   0.0018

Quantifying heterogeneity:
 tau^2 = 1.2333; H = 3.91 [3.15; 4.84]; I^2 = 93.4% [89.9%; 95.7%]

Test of heterogeneity:
      Q d.f.  p-value
 137.34    9 < 0.0001

Details on meta-analytical method:
- Mantel-Haenszel method (without continuity correction)
- DerSimonian-Laird estimator for tau^2
- Continuity correction of 0.5 in studies with zero cell frequencies
```

Table 3.2: Output presenting fixed and random effects analysis for Odds ratio using Mantel-Haenszel method.

As one can see from the results of Table 3.2, the value of Odds Ratio is large (3.25 and 3.35 for the fixed and random effects estimate respectively). The difference between fixed and random effects models is not that large. However, the random effects estimate has a lot larger variability depicted in to much larger CIs (1.57, 7.15), still however statistically different than 1. All heterogeneity indeces indicate a problem of heterogeneity between trials (H=3.91, I^2=93.4%, Q=137.34, p<0.001).

The forest plot (Table 3.3) indicates that the majority of Odds ratio values are greater than 1 and that trials 5,8,9 have greater variability than other trials. Heterogeneity index I^2=93% , p<0.01 shows a problem of heterogeneity between trials. The funnel plot (Figure 3.1) indicates some publication bias for six studies (1,2,4,6,7,10). These studies present unusually low variability according to the distance from the common value of Odds ratio.

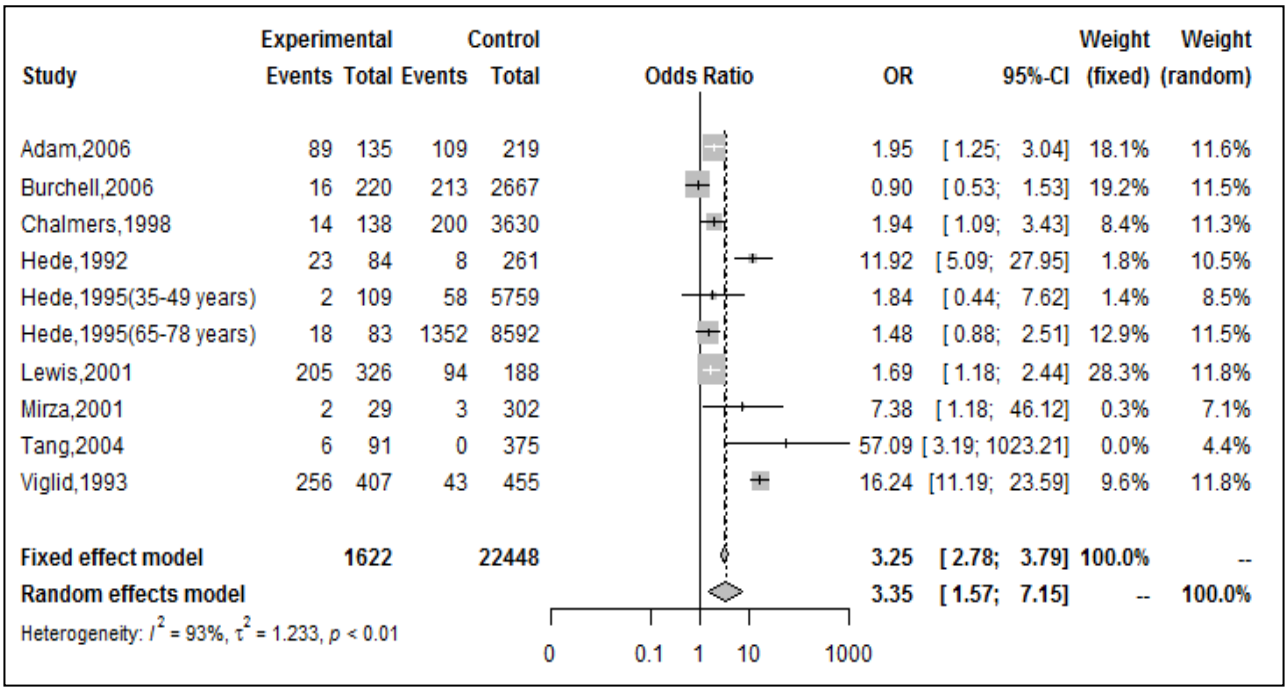

Table 3.3: Forest plot of fixed and random effects analysis for Odds ratio using Mantel-Haenszel method.

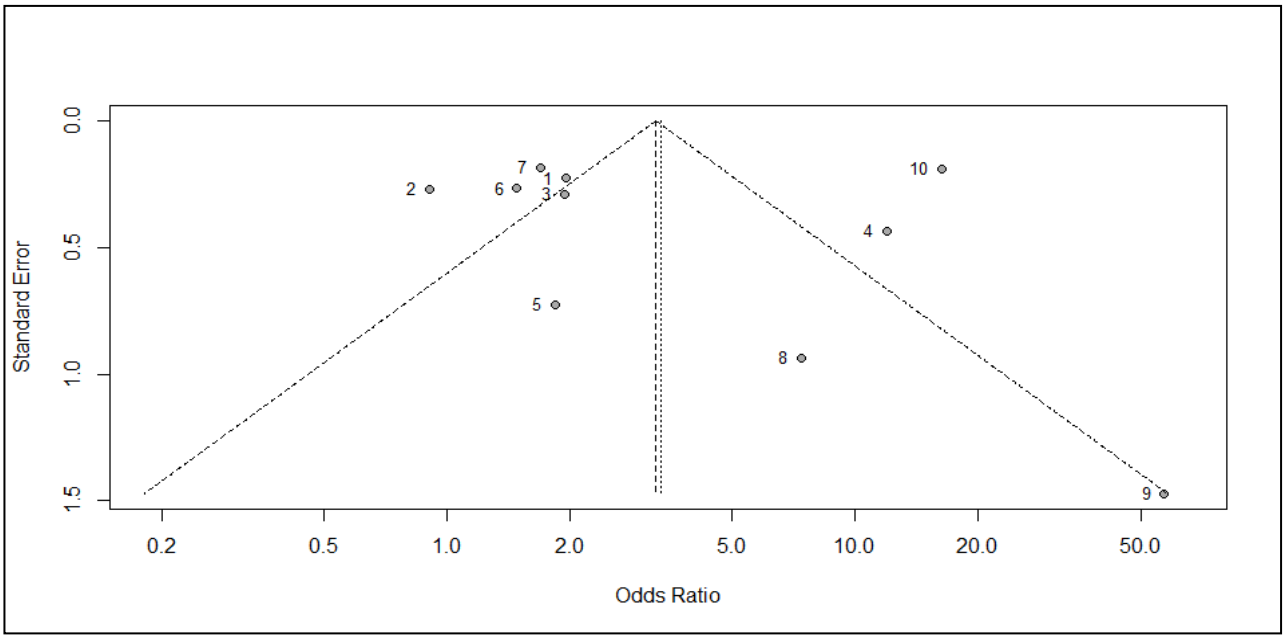

Figure 3.2: Funnel plot of Odds ratio using Mantel-Haenszel method.

Next, results of Risk ratio using Mantel-Haenszel method are presented:

mRR<-

metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_sbg1$Control.events,data_sbg1$Control.total,sm="RR" , method ="MH" , MH.exact = TRUE)

mRR

forest(mRR,studlab=study)

```
> mRR
         RR              95%-CI %w(fixed) %w(random)
1    1.3246 [1.1063;   1.5858]      25.8        13.0
2    0.9106 [0.5584;   1.4851]      10.1        11.9
3    1.8413 [1.1007;   3.0801]       4.5        11.8
4    8.9330 [4.1527;  19.2162]       1.2        10.4
5    1.8219 [0.4507;   7.3645]       0.7         6.9
6    1.3782 [0.9131;   2.0803]       8.0        12.2
7    1.2577 [1.0658;   1.4840]      37.0        13.0
8    6.9425 [1.2087;  39.8777]       0.2         5.4
9   53.3497 [3.0329; 938.4402]       0.0         2.7
10   6.6556 [4.9601;   8.9308]      12.6        12.7

Number of studies combined: k = 10

                         RR            95%-CI     z  p-value
Fixed effect model   2.0914 [1.8853; 2.3201] 13.94 < 0.0001
Random effects model 2.4303 [1.4307; 4.1280]  3.29   0.0010

Quantifying heterogeneity:
 tau^2 = 0.5529; H = 4.17 [3.39; 5.12]; I^2 = 94.2% [91.3%; 96.2%]

Test of heterogeneity:
      Q d.f.  p-value
 156.38    9 < 0.0001
```

Table 3.4: Output presenting fixed and random effects analysis for Risk ratio using Mantel-Haenszel method.

As one can see from the results of Table 3.3, the value of Risk Ratio is 2.09 and 2.43 for the fixed and random effects estimate respectively. The difference between fixed and random effects models is not that large. However, the random effects estimate has a lot larger variability depicted in to much larger CIs (1.43, 4.13), still however statistically different than 1. All heterogeneity indeces indicate a problem of heterogeneity between trials (H=4.17, I^2=94.2%, Q=156.38, p<0.001). The forest plot indicates that the majority of Risk ratio values are greater than 1 and that trials 5,8,9 have greater variability than other trials. Heterogeneity index I^2=94% , p<0.01 shows a problem of heterogeneity between trials. The funnel plot (Figure 3.2) indicates some

publication bias for seven studies (1,2,4,6,7,9,10). These studies present unusually low variability according to the distance from the common value of Risk ratio.

Study	Experimental Events	Experimental Total	Control Events	Control Total	Risk Ratio	RR	95%-CI	Weight (fixed)	Weight (random)
Adam,2006	89	135	109	219		1.32	[1.11; 1.59]	25.8%	13.0%
Burchell,2006	16	220	213	2667		0.91	[0.56; 1.49]	10.1%	11.9%
Chalmers,1998	14	138	200	3630		1.84	[1.10; 3.08]	4.5%	11.8%
Hede,1992	23	84	8	261		8.93	[4.15; 19.22]	1.2%	10.4%
Hede,1995(35-49years)	2	109	58	5759		1.82	[0.45; 7.36]	0.7%	6.9%
Hede,1995(65-78 years)	18	83	1352	8592		1.38	[0.91; 2.08]	8.0%	12.2%
Lewis,2001	205	326	94	188		1.26	[1.07; 1.48]	37.0%	13.0%
Mirza,2001	2	29	3	302		6.94	[1.21; 39.88]	0.2%	5.4%
Tang,2004	6	91	0	375		53.35	[3.03; 938.44]	0.0%	2.7%
Viglid,1993	256	407	43	455		6.66	[4.96; 8.93]	12.6%	12.7%
Fixed effect model		**1622**		**22448**		**2.09**	**[1.89; 2.32]**	**100.0%**	--
Random effects model						**2.43**	**[1.43; 4.13]**	--	**100.0%**

Heterogeneity: I^2 = 94%, τ^2 = 0.5529, p < 0.01

0.01 0.1 1 10 100

Table 3.5: Forest plot presenting random effects analysis for Risk ratio using Mantel-Haenszel method.

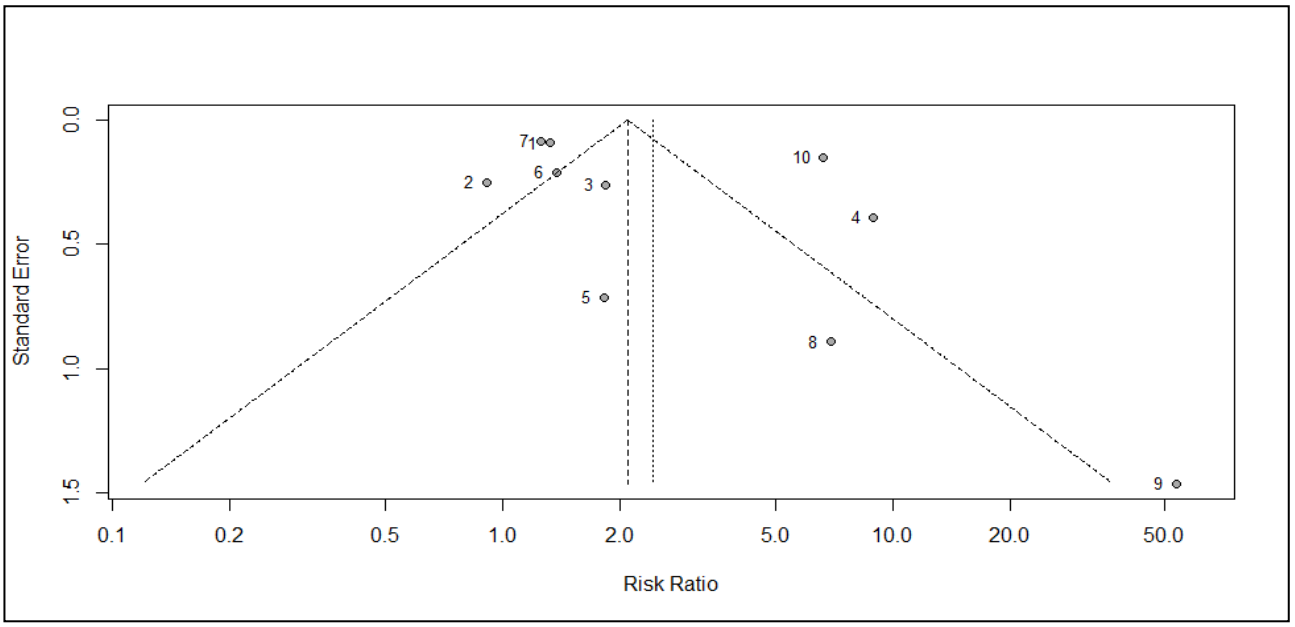

Figure 3.3: Funnel plot of Risk ratio using Mantel-Haenszel method.

Next, results of Risk difference using Mantel-Haenszel method are presented:

mRD<-

metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_sbg1$Control.events,data_sbg1$Control.total,sm="RD" , method ="MH" , MH.exact = TRUE)

mRD

forest(mRD,studlab=study)

```
> mRD
         RD             95%-CI %W(fixed) %W(random)
1    0.1615 [ 0.0577; 0.2654]       7.5        9.6
2   -0.0071 [-0.0430; 0.0287]      18.4       10.3
3    0.0464 [-0.0046; 0.0973]      12.0       10.2
4    0.2432 [ 0.1455; 0.3408]       5.7        9.7
5    0.0083 [-0.0170; 0.0336]       9.7       10.3
6    0.0595 [-0.0295; 0.1485]       7.4        9.8
7    0.1288 [ 0.0402; 0.2175]      10.8        9.8
8    0.0590 [-0.0339; 0.1519]       2.4        9.8
9    0.0659 [ 0.0134; 0.1184]       6.6       10.2
10   0.5345 [ 0.4804; 0.5886]      19.4       10.2

Number of studies combined: k = 10

                         RD            95%-CI     z  p-value
Fixed effect model   0.1591 [0.1381; 0.1801] 14.84 < 0.0001
Random effects model 0.1295 [0.0086; 0.2503]  2.10   0.0357

Quantifying heterogeneity:
 tau^2 = 0.0366; H = 7.05 [6.08; 8.18]; I^2 = 98.0% [97.3%; 98.5%]

Test of heterogeneity:
      Q d.f.  p-value
 447.55    9 < 0.0001
```

Table 3.6: Output presenting fixed and random effects analysis for Risk difference using Mantel-Haenszel method.

As one can see from the results of Table 3.6, the value of Risk Difference is 0.16 and 0.13 for the fixed and random effects estimate respectively. The difference between fixed and random effects models is not that large. However, the random effects estimate has a lot larger variability depicted in to much larger CIs (0.01, 0.25), still however statistically different than 0. All heterogeneity indeces indicate a problem of heterogeneity between trials (H=7.05, I^2=98%, Q=447.55, p<0.001). The forest plot indicates that the majority of Risk differnces values are greater than 0 and that all trials contribute practically the same information for the calculation of the common risk difference. Heterogeneity index I^2=98% , p<0.01 shows a problem of

heterogeneity between trials. The funnel plot (Figure 3.3) indicates some publication bias for seven studies (2,3,5,6,8,9,10). These studies present unusually low variability according to the distance from the common value of Risk difference.

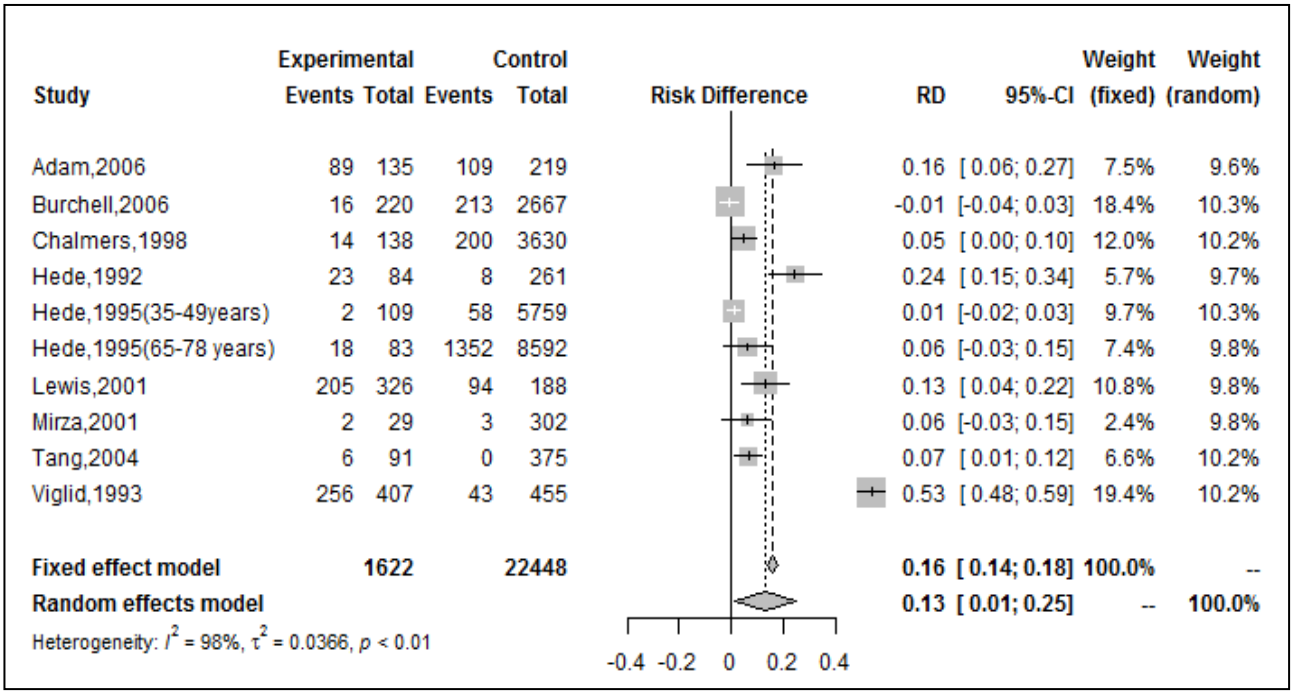

Table 3.7: Forest plot presenting random effects analysis for Risk difference using Mantel-Haenszel method.

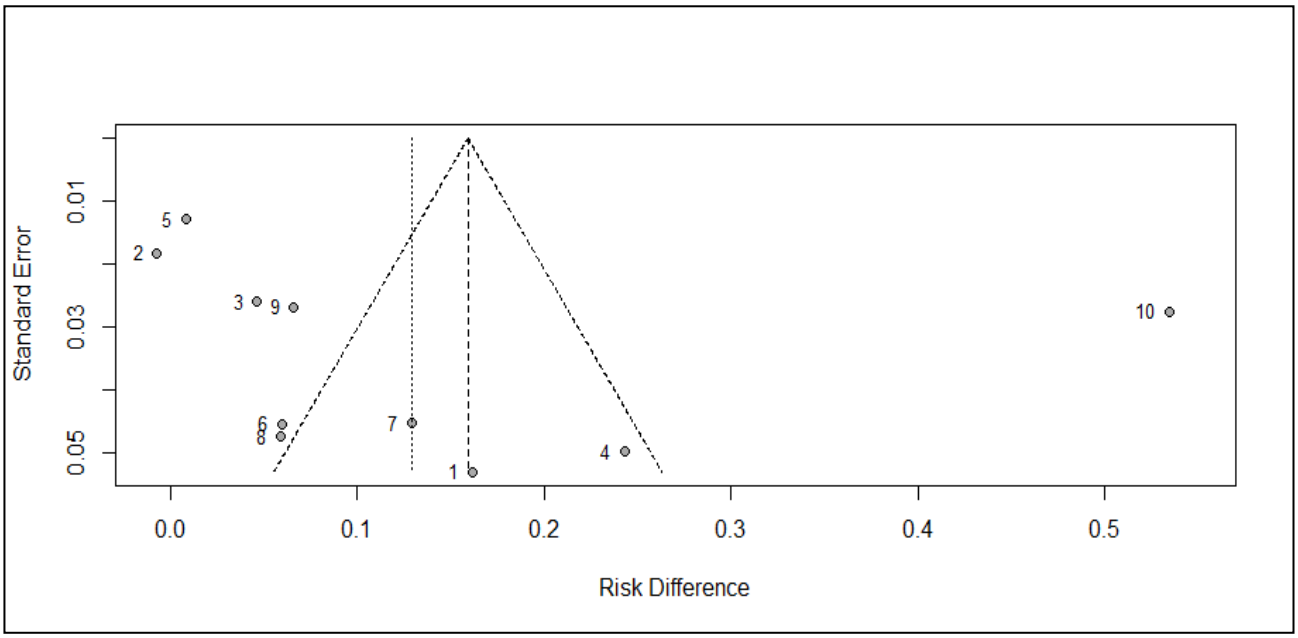

Figure 3.4: Funnel plot of Risk ratio using Mantel-Haenszel method

Next, results of Odds ratio using Peto method are presented:

pOR<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_sbg1$Control.events,data_sbg1$Control.total,sm="OR" , method ="Peto" , MH.exact = TRUE)
pOR
forest(pOR,studlab=study)

```
> pOR
          OR               95%-CI %W(fixed) %W(random)
1     1.9223 [ 1.2488;     2.9592]      14.9        12.0
2     0.9069 [ 0.5453;     1.5083]      10.7        11.8
3     2.3752 [ 1.1397;     4.9501]       5.1        11.2
4    19.3881 [ 8.2166;    45.7487]       3.8        10.8
5     2.2654 [ 0.3445;    14.8988]       0.8         7.2
6     1.5643 [ 0.8648;     2.8298]       7.9        11.6
7     1.6963 [ 1.1793;     2.4399]      20.9        12.1
8    52.2401 [ 2.3087; 1182.0492]       0.3         4.2
9   177.0692 [23.2692; 1347.4241]       0.7         6.8
10   10.5540 [ 7.9703;    13.9752]      35.1        12.2

Number of studies combined: k = 10

                         OR           95%-CI     z  p-value
Fixed effect model   3.5486 [3.0051; 4.1905] 14.93 < 0.0001
Random effects model 4.3263 [1.9763; 9.4705]  3.66   0.0002

Quantifying heterogeneity:
 tau^2 = 1.2848; H = 4.08 [3.31; 5.03]; I^2 = 94.0% [90.9%; 96.0%]

Test of heterogeneity:
      Q d.f.  p-value
 149.96    9 < 0.0001
```

Table 3.8: Output presenting fixed and random effects analysis for Odds ratio using Peto method.

As one can see from the results of Table 3.8, the value of Odds Ratio is 3.54 and 4.32 for the fixed and random effects estimate respectively. The difference between fixed and random effects models is large. However, the random effects estimate has a lot larger variability depicted into much larger CIs (1.98, 9.47), still however statistically different than 1. All heterogeneity indeces indicate a problem of heterogeneity between trials (H=4.08, I^2=94%, Q=149.96, p<0.001).

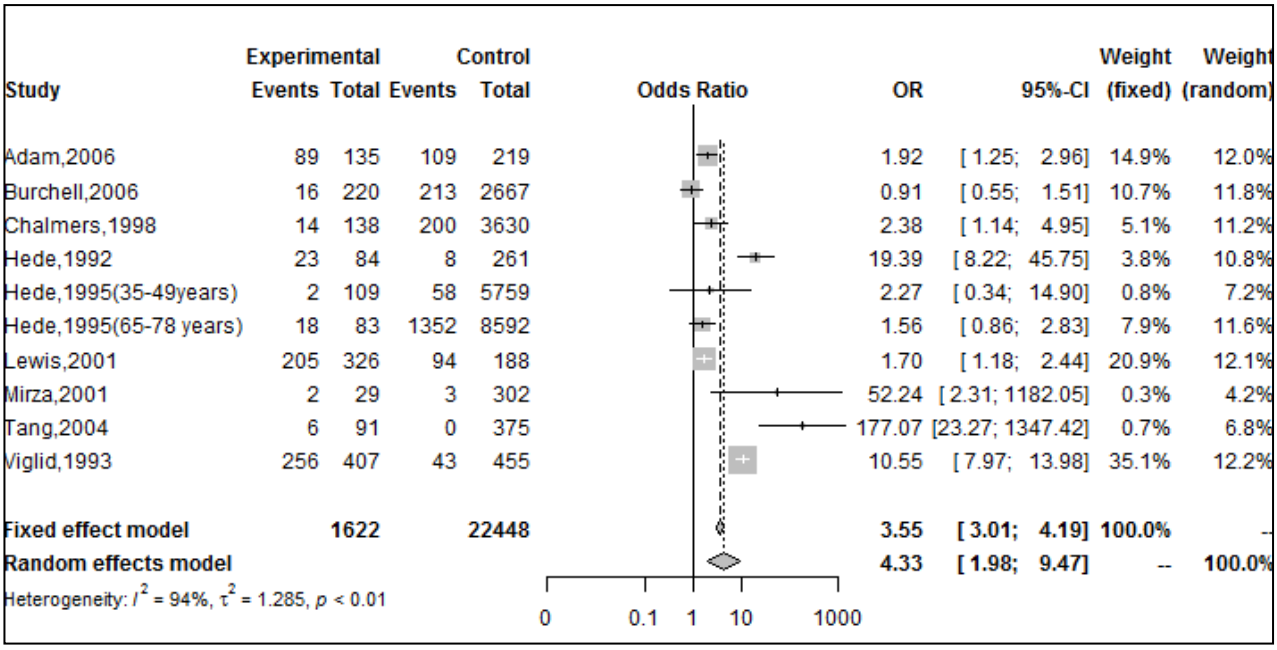

Table 3.9: Forest plot presenting random effects analysis for Odds ratio using Peto method.

The forest plot indicates that the majority of Odds ratio values are greater than 1 and that trials 5,8,9 have greater variability than other trials. Heterogeneity index I^2=94% , p<0.01 shows a problem of heterogeneity between trials. The funnel plot (Figure 3.4) indicates some publication bias for studies 1,2,4,6,7,9,10. These studies present unusually low variability according to the distance from the common value of Odds ratio.

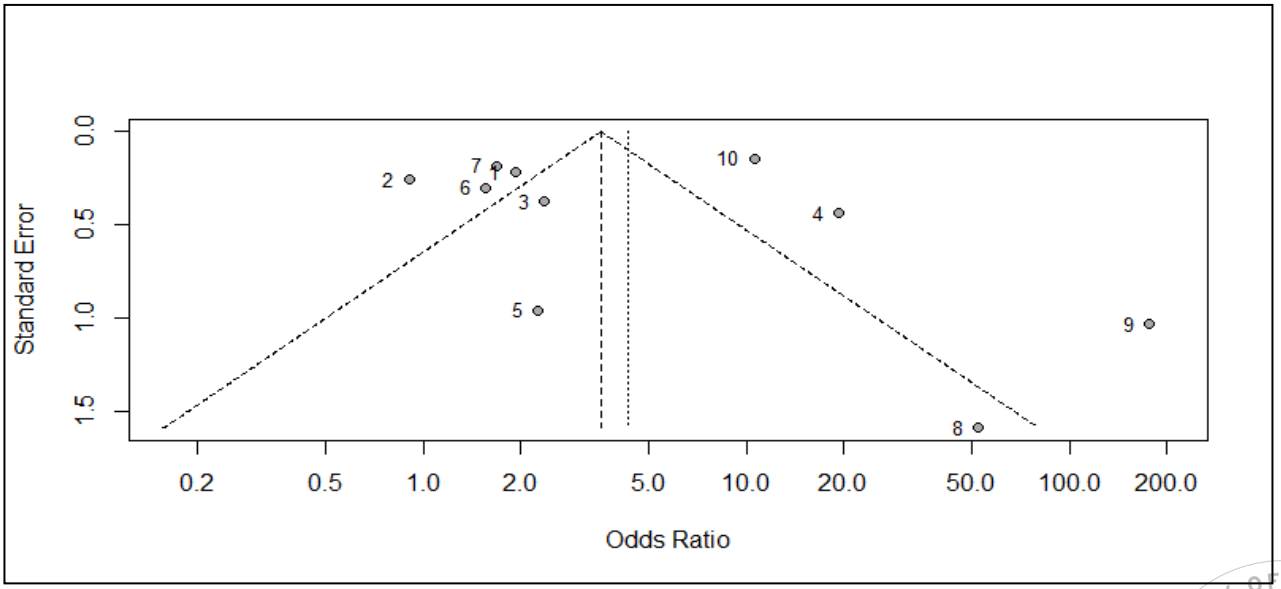

Figure 3.5: Funnel plot of Odds ratio using Peto method.

Next, results of Odds ratio using Inverse varianse method are presented:

ivOR<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_sbg1$Control.events,data_sbg1$Control.total,sm="OR" , method ="Inverse" , MH.exact = TRUE)
ivOR
forest(ivOR,studlab=study)

```
> ivOR
         OR                 95%-CI %w(fixed) %w(random)
1    1.9525 [ 1.2529;     3.0428]       15.7        11.7
2    0.9036 [ 0.5331;     1.5317]       11.1        11.5
3    1.9363 [ 1.0943;     3.4263]        9.5        11.4
4   11.9242 [ 5.0880;    27.9453]        4.3        10.5
5    1.8373 [ 0.4430;     7.6201]        1.5         8.5
6    1.4829 [ 0.8770;     2.5075]       11.2        11.5
7    1.6942 [ 1.1777;     2.4372]       23.3        11.8
8    7.3827 [ 1.1819;    46.1166]        0.9         7.1
9   57.0936 [ 3.1857; 1023.2145]        0.4         4.4
10 16.2440 [11.1870;    23.5869]       22.2        11.8

Number of studies combined: k = 10

                           OR           95%-CI     z  p-value
Fixed effect model   2.9737 [2.4947; 3.5448] 12.16 < 0.0001
Random effects model 3.3487 [1.5731; 7.1285]  3.14   0.0017

Quantifying heterogeneity:
 tau^2 = 1.2240; H = 3.89 [3.14; 4.83]; I^2 = 93.4% [89.9%; 95.7%]

Test of heterogeneity:
      Q d.f.  p-value
 136.37    9 < 0.0001
```

Table 3.10: Output presenting fixed and random effects analysis for Odds ratio using Inverse variance method.

As one can see from the results of Table 3.10, the value of Odds ratio is 2.97 and 3.34 for the fixed and random effects estimate respectively. The difference between fixed and random effects models is large. However, the random effects estimate has a lot larger variability depicted into much larger CIs (1.57, 7.13), still however statistically different than 1. All heterogeneity indeces indicate a problem of heterogeneity between trials (H=3.89, I^2=93.4%, Q=136.37, p<0.001).

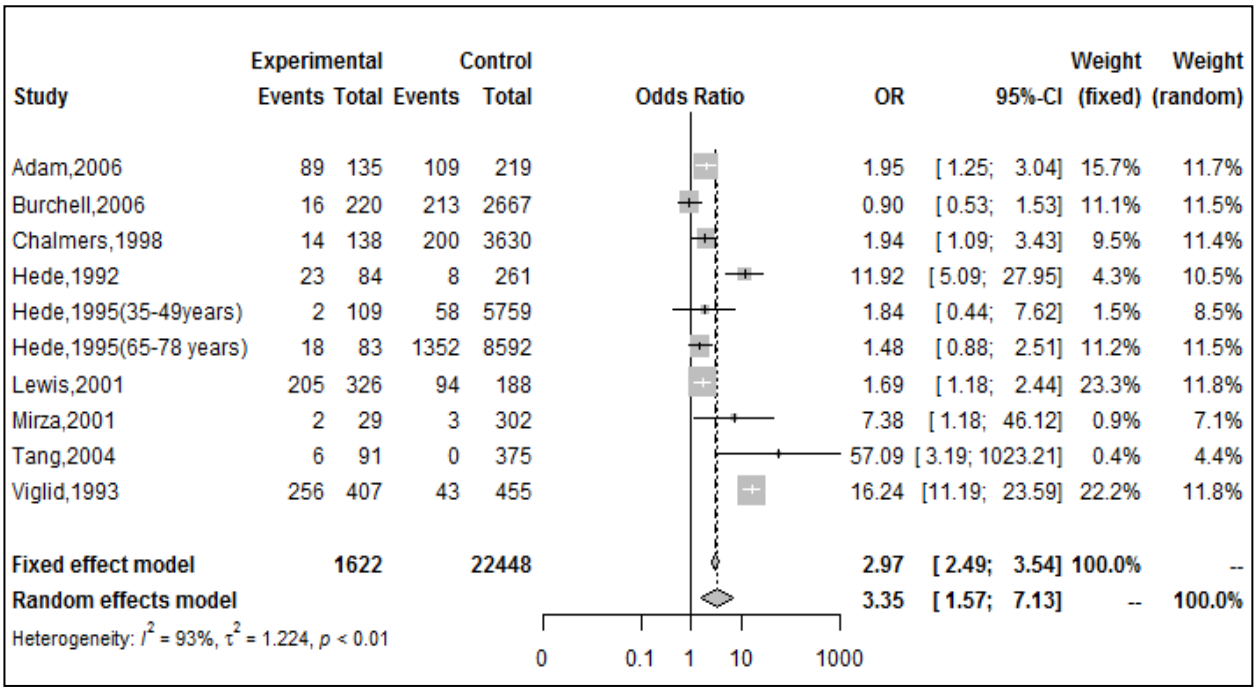

Table 3.11: Forest plot presenting random effects analysis for Odds ratio using Inverse variance method.

The forest plot indicates that the majority of Odds ratio values are greater than 1 and that and that trials 5,8,9 have greater variability than other trials. Heterogeneity index I^2=93% , p<0.01 shows a problem of heterogeneity between trials. The funnel plot (Figure 3.5) indicates some publication bias for studies 2,4,6,7,9,10. These studies present unusually low variability according to the distance from the common value of Odds Ratio.

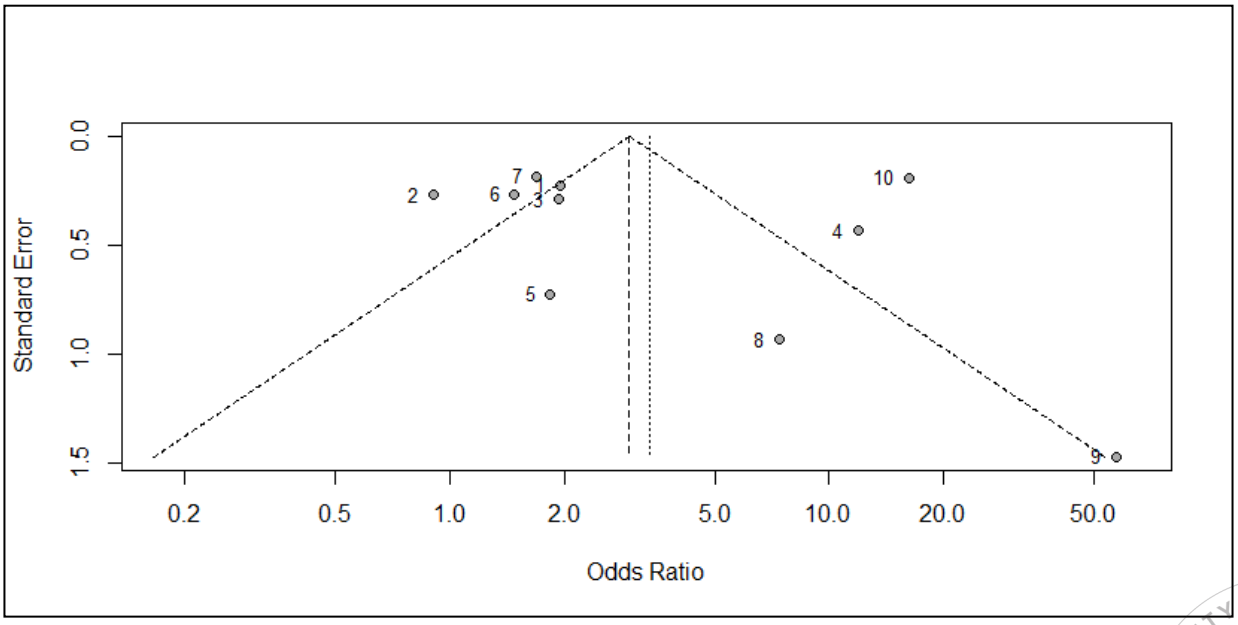

Figure 3.6: Funnel plot of Odds ratio using Inverse variance method.

Next, results of Risk ratio using Inverse variance method are presented:

ivRR<-

metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_sbg1$Control.events,data_sbg1$Control.total,sm="RR" , method ="Inverse" , MH.exact = TRUE)

ivRR

forest(ivRR,studlab=study)

```
> ivRR
         RR              95%-CI %W(fixed) %W(random)
1    1.3246 [1.1063;   1.5858]      32.3       13.3
2    0.9106 [0.5584;   1.4851]       4.4       12.0
3    1.8413 [1.1007;   3.0801]       4.0       11.8
4    8.9330 [4.1527;  19.2162]       1.8       10.2
5    1.8219 [0.4507;   7.3645]       0.5        6.5
6    1.3782 [0.9131;   2.0803]       6.2       12.4
7    1.2577 [1.0658;   1.4840]      38.2       13.3
8    6.9425 [1.2087;  39.8777]       0.3        5.1
9   53.3497 [3.0329; 938.4402]       0.1        2.5
10   6.6556 [4.9601;   8.9308]      12.1       12.9

Number of studies combined: k = 10

                         RR            95%-CI    z  p-value
Fixed effect model   1.6522 [1.4915; 1.8303] 9.62 < 0.0001
Random effects model 2.3940 [1.4556; 3.9374] 3.44   0.0006

Quantifying heterogeneity:
 tau^2 = 0.4765; H = 3.89 [3.13; 4.82]; I^2 = 93.4% [89.8%; 95.7%]

Test of heterogeneity:
      Q d.f.  p-value
 136.00    9 < 0.0001
```

Table 3.12: Output presenting fixed and random effects analysis for Risk ratio using Inverse variance method.

As one can see from the results of Table 3.12, the value of Risk ratio is 1.65 and 2.4 for the fixed and random effects estimate respectively. The difference between fixed and random effects models is large. However, the random effects estimate has a lot larger variability depicted into much larger CIs (1.45, 3.94), still however statistically different than 1. All heterogeneity indeces indicate a problem of heterogeneity between trials (H=3.89, I^2=93.4%, Q=136, p<0.001).

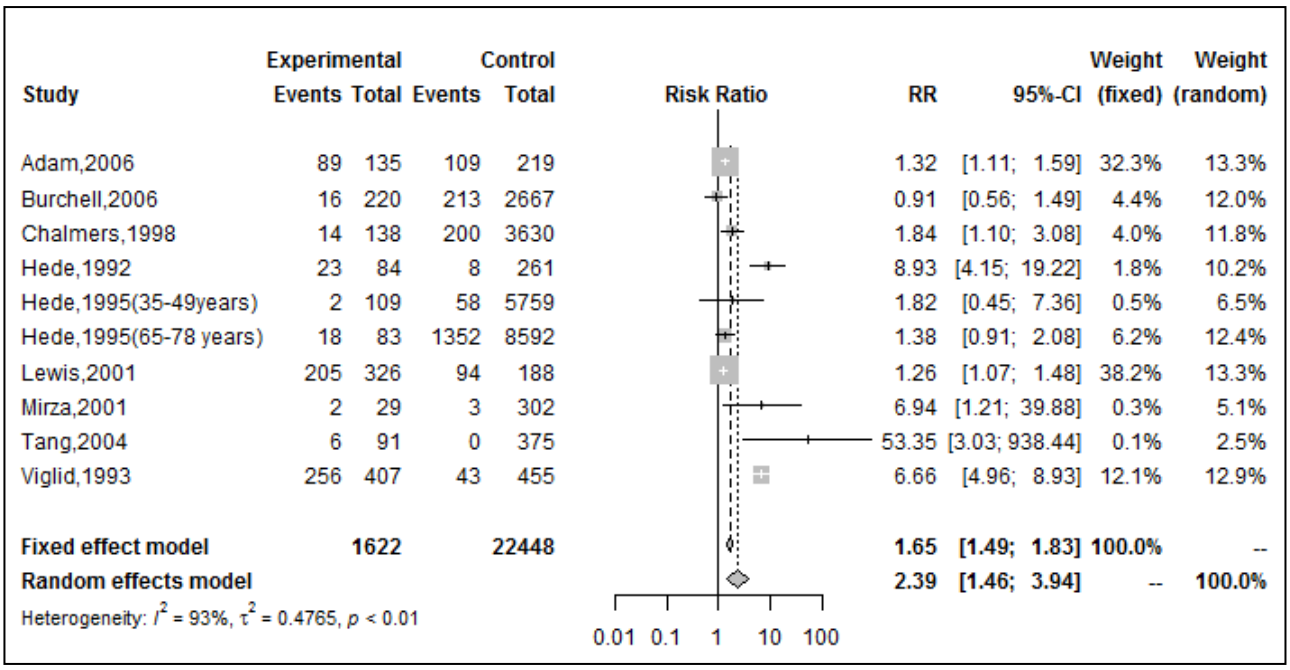

Table 3.13: Forest plot of random effects analysis for Risk ratio using Inverse variance method.

The forest plot indicates that the majority of Risk ratio values are greater than 1 and that trials 5,8,9 have greater variability than other trials. Heterogeneity index I^2=93% , p<0.01 shows a problem of heterogeneity between trials. The funnel plot (Figure 3.6) indicates some publication bias for studies . These studies present unusually low variability according to the distance from the common value of Risk ratio.

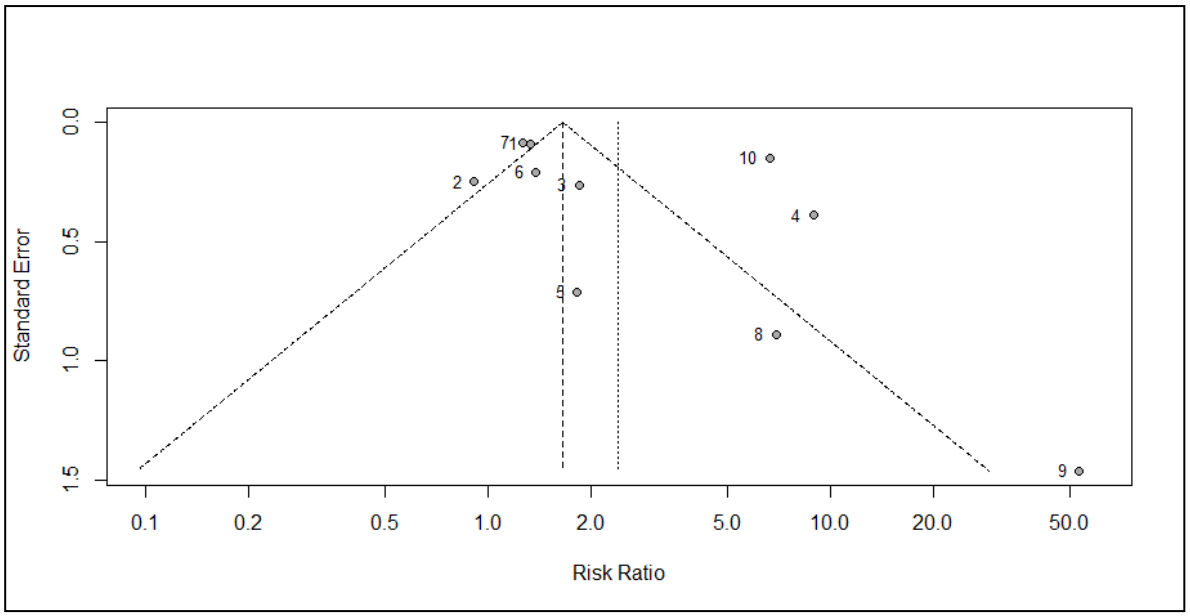

Figure 3.7: Funnel plot of Risk ratio using Inverse variance method.

Next, result of Risk difference using Inverse variance method are presented:

ivRD<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_sbg1$Control.events,data_sbg1$Control.total,sm="RD" , method ="Inverse" , MH.exact = TRUE)
ivRD
forest(ivRD,studlab=study)

```
> ivRD
          RD             95%-CI %w(fixed) %w(random)
1    0.1615 [ 0.0577; 0.2654]        2.3         9.5
2   -0.0071 [-0.0430; 0.0287]       19.5        10.4
3    0.0464 [-0.0046; 0.0973]        9.7        10.3
4    0.2432 [ 0.1455; 0.3408]        2.6         9.6
5    0.0083 [-0.0170; 0.0336]       39.0        10.4
6    0.0595 [-0.0295; 0.1485]        3.2         9.8
7    0.1288 [ 0.0402; 0.2175]        3.2         9.8
8    0.0590 [-0.0339; 0.1519]        2.9         9.7
9    0.0659 [ 0.0134; 0.1184]        9.1        10.2
10   0.5345 [ 0.4804; 0.5886]        8.6        10.2

Number of studies combined: k = 10

                         RD            95%-CI     z  p-value
Fixed effect model   0.0759 [0.0600; 0.0917] 9.40 < 0.0001
Random effects model 0.1293 [0.0235; 0.2350] 2.40   0.0166

Quantifying heterogeneity:
 tau^2 = 0.0277; H = 6.16 [5.24; 7.24]; I^2 = 97.4% [96.4%; 98.1%]

Test of heterogeneity:
      Q d.f.  p-value
 341.16    9 < 0.0001
```

Table 3.14: Output presenting fixed and random effects analysis for Risk differenceusing Inverse variance method.

As one can see from the results of Table 3.14, the value of Risk ratio is 0.08 and 0.13 for the fixed and random effects estimate respectively. The difference between fixed and random effects models is large. However, the random effects estimate has a lot larger variability depicted into much larger CIs (0.02, 0.24), still however statistically different than 0. All heterogeneity indeces indicate a problem of heterogeneity between trials (H=6.19, I^2=97.4%, Q=341.16, p<0.001).

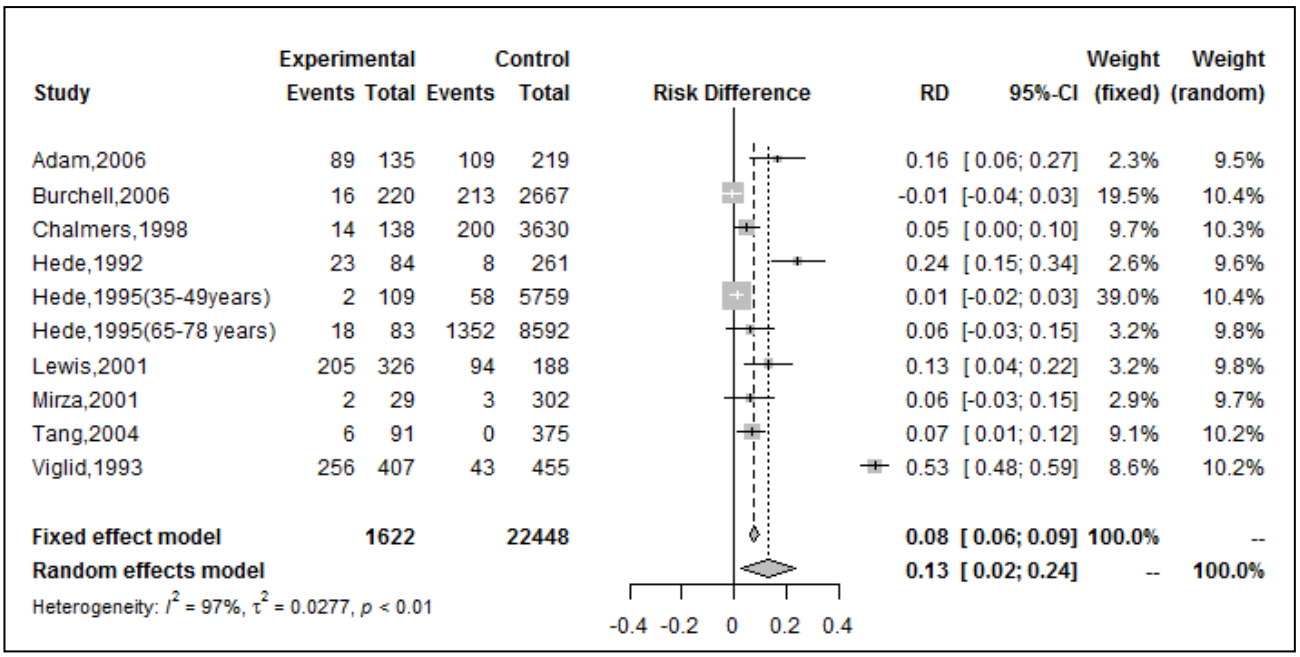

Table 3.15: Forest plot of random effects analysis for Odds ratio using Inverse variance method.

The forest plot indicates that the majority of Risk differences values are greater than 0 and that all trials contribute practically the same information for the calculation of the common risk difference. Heterogeneity index I^2=97% , p<0.01 shows a problem of heterogeneity between trials. The funnel plot (Figure 3.7) indicates some publication bias for studies 2,4,5,10. These studies present unusually low variability according to the distance from the common value of Risk difference.

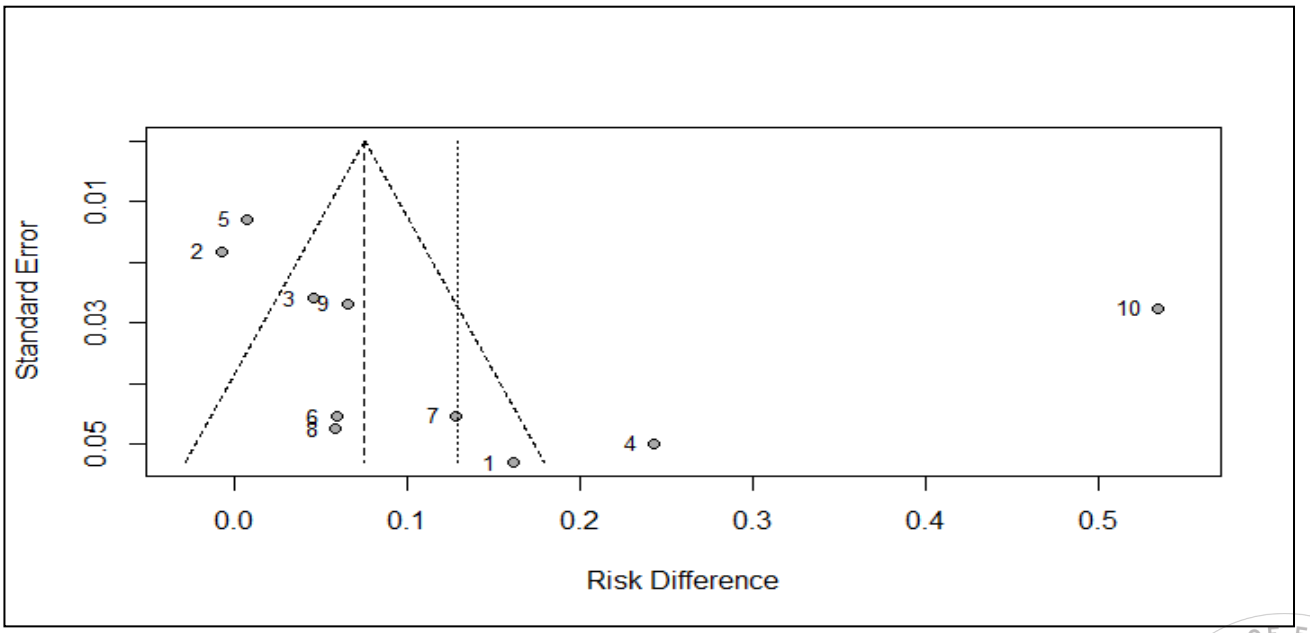

Figure 3.8: Funnel plot of Risk difference using Inverse variance method.

We had also five other measurements for which we investigated results. We assessed the number of decayed, missing and filled dental surfaces or teeth; both these indices are expressed as a continuous variable. The number of decayed, missing and filled teeth reflects a person's lifetime experience of dental caries. This is because both dental decay and its treatment leave permanent marks, either through the presence of fillings or the loss of affected teeth by extraction. The total number of teeth (T) and surfaces (S) that are decayed (D), missing because of pathology (M) or filled (F) are measures referred to as DMFT and DMFS respectively (Kisely, et. al.,2011). In both, an increase in score means greater dental decay. Decayed surfaces, DMFS, decayed teeth, missing teeth, DMFT. For all these measurements we had all the information needed from each study so we can do the meta-analysis. We had the mean, the standard deviation and the total number for the control and the psychiatric group as well.

Kisely et. al., (2011) defined the following clarifications on the outcome measures:

The primary outcome of this study was edentulousness, usually expressed as a dichotomous variable.

We also assessed the number of decayed, missing and filled dental surfaces or teeth; both these indices are expressed as a continuous variable. The number of decayed, missing and filled teeth reflects a person's lifetime experience of dental caries. This is because both dental decay and its treatment leave permanent marks, either through the presence of fillings or the loss of affected teeth by extraction. The total number of teeth (T) and surfaces (S) that are decayed, missing because of pathology (M) or filled (F) are measures referred to as DMFT and DMFS respectively. In both, an increase in score means greater dental decay.

Scores for DMFT and DMFS vary widely by country, from mean DMFT scores of under 5 in India to 12.8 in the most recent community survey in a high income country (Australia).16–18 Scores for DMFS are higher than for DMFT as the former counts damage to each surface of each tooth rather than counting the tooth as a single unit; anterior teeth have four surfaces and posterior teeth five. In interpreting both, it is useful to recall that humans have 32 permanent teeth. The maximum possible DMFT score is therefore 32, whereas the maximum DMFS is 148.

So we continued the meta-analysis for each measurement with the random effects method. Decayed surfaces is the first measurement. The following R commands were used for the analysis:

m1<-c(3.1,5.9,6.4)

mc1<-c(0.9,1.5,2.07)

sd1<-c(11.6,8.3,6.67)

n1<-c(109,23,37)

sdc1<-c(0.1,0.3,2.53)

nco1<-c(762,353,29)

ma1<- rma(yi=m1-mc1,vi=sd1^2/n1+sdc1^2/nco1,method = "DL")

ma1

study1<-c('Hede, 1995(35–49 years)','Hede, 1995 (65–78 years)','Stoefe; 1990')

forest(ma1,slab=study1,xlab='Decayed surfaces')

```
> ma1

Random-Effects Model (k = 3; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.0928 (SE = 1.8120)
tau (square root of estimated tau^2 value):      0.3046
I^2 (total heterogeneity / total variability):   5.12%
H^2 (total variability / sampling variability):  1.05

Test for Heterogeneity:
Q(df = 2) = 2.1079, p-val = 0.3486

Model Results:

estimate       se     zval     pval    ci.lb    ci.ub
  3.4191   0.7588   4.5061   <.0001   1.9320   4.9063      ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3.16: Output presenting random effects analysis for mean difference of decayed surfaces between mentally diseased and controls.

The results of Table 3.16 indicate that heterogeneity is absent (Q=2.11, p=0.349) and there is statistically significant difference (mean difference=3.42, p<0.001) in Decayed surfaces measurement between people with mental disease and control (3.17). . The funnel plot (Figure 3.8) does not indicate any significant publication

bias for the three studies. These studies present no variability according to the distance from the common value of mean difference.

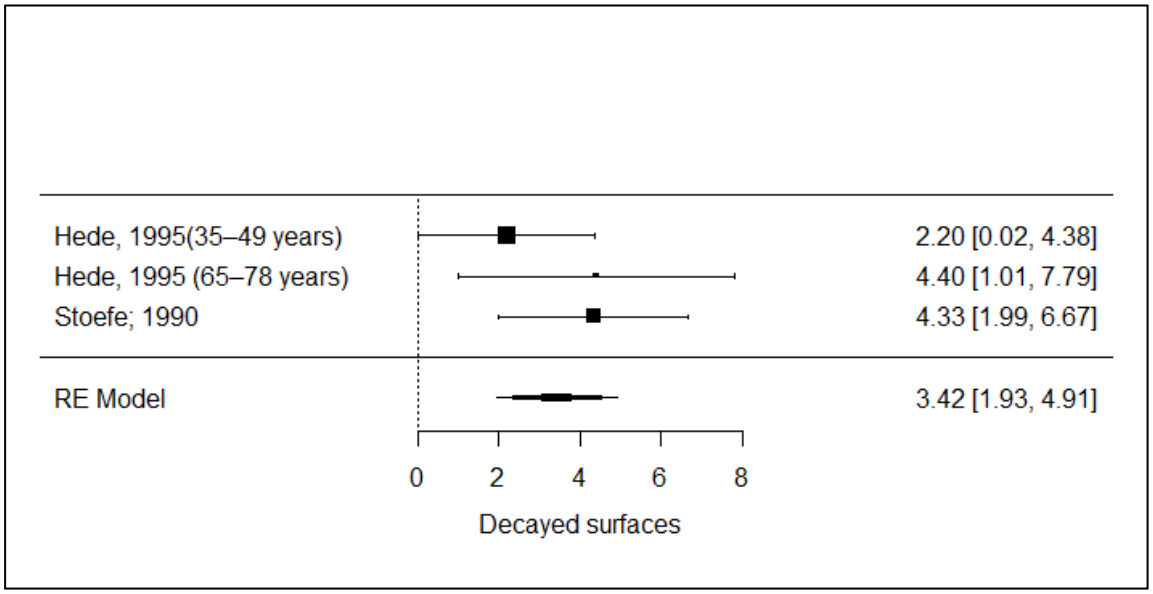

Table 3.17: Forest plot of random effects analysis for mean difference of decayed surfaces between mentally diseased and controls.

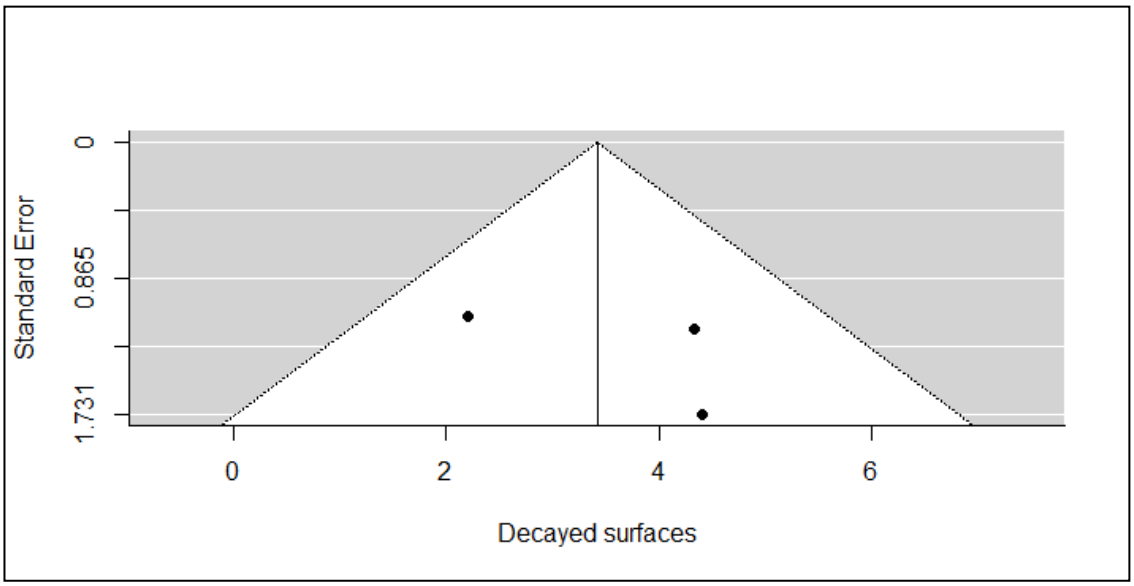

Figure 3.9: Funnel plot of mean difference of Decayed Surfaces.

Next measurement under examination is the DMFS. The following R commands were used for the analysis.

```
m2<-c(68.3,120.2,31.9)
mc2<-c(46.6,104.1,27.4)
sd2<-c(33.3,27.8,22)
sdc2<-c(0.7,1.7,20)
ma2<- rma(yi=m2-mc2,vi=sd2^2/n1+sdc2^2/nco1,method = "DL")
ma2
study2<-c('Hede, 1995(35–49 years)','Hede, 1995 (65–78 years)','Stiefel; 1990')
forest(ma2,slab=study2,xlab='DMFS')
```

DMFS is the second measurement. The following R commands were used for the analysis.

```
> ma2

Random-Effects Model (k = 3; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 64.5202 (SE = 88.1040)
tau (square root of estimated tau^2 value):      8.0324
I^2 (total heterogeneity / total variability):   75.01%
H^2 (total variability / sampling variability):  4.00

Test for Heterogeneity:
Q(df = 2) = 8.0040, p-val = 0.0183

Model Results:

estimate       se     zval     pval    ci.lb    ci.ub
 14.5951   5.3823   2.7117   0.0067   4.0461  25.1441       **

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3.18: Output presenting random effects analysis for mean score of DMFS between mentally diseased and controls.

The results of Table 3.18 indicate that there is some heterogeneity between studies (Q=8.00, p=0.018) and there is statistically significant difference (mean difference=14.60, p<0.007) in DMFS measurement between people with mental disease and control (3.19). The funnel plot (Figure 3.9) does not indicate any

significant publication bias for the three studies. These studies present no variability according to the distance from the common value of mean difference.

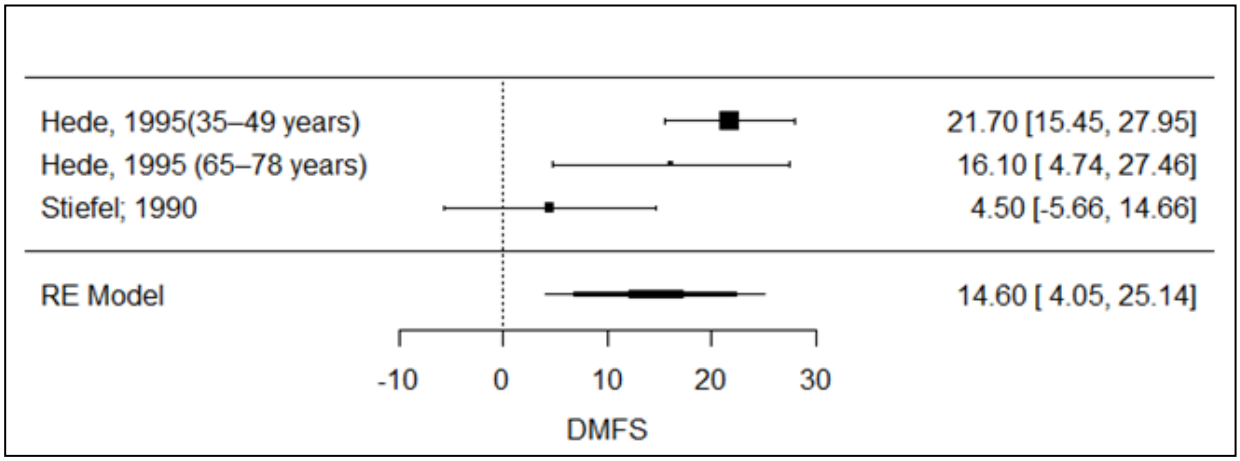

Table 3.19: Forest plot of random effects analysis for mean difference of decayed surfaces between mentally diseased and controls.

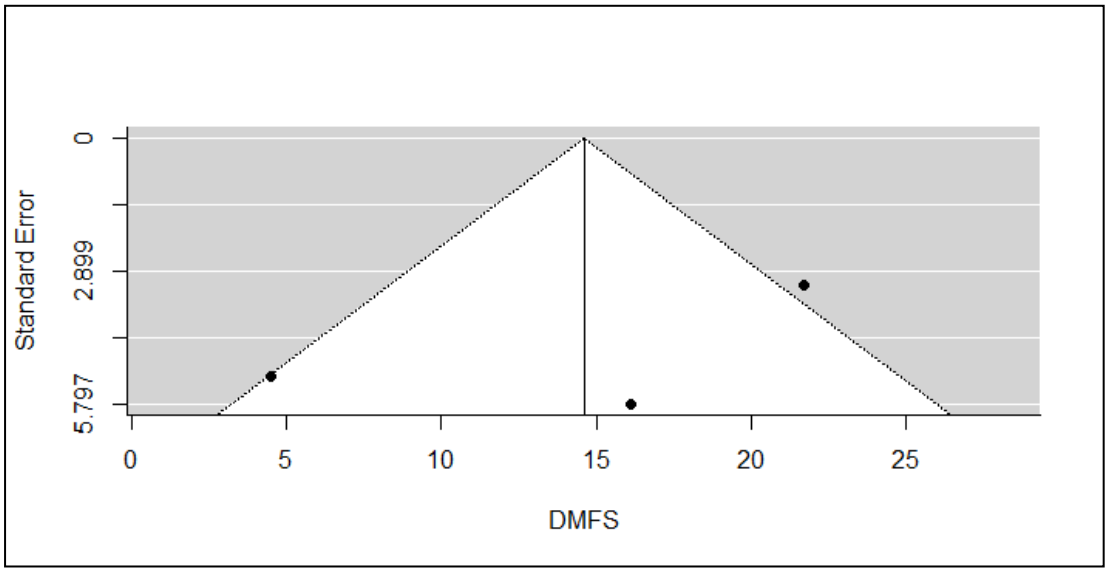

Figure 3.10: Funnel plot of mean difference of DMFS.

Next measurement under examination is the number of decayed teeth. The following R commands were used for the analysis.

m3<-c(9.16,7.95)

mc3<-c(2.55,2.9)

sd3<-c(5.2,6.86)

sdc3<-c(3.12,2.19)

n3<-c(54,565)

nco3<-c(7139,261)

ma3<- rma(yi=m3-mc3,vi=sd3^2/n3+sdc3^2/nco3, method = "DL")

ma3

study3<-c('Ramon, 2003 (18–34 years)','Velasco, 1997')

forest(ma3,slab=study3,xlab='Decayed Teeth')

```
> ma3

Random-Effects Model (k = 2; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 0.9149 (SE = 1.7208)
tau (square root of estimated tau^2 value):      0.9565
I^2 (total heterogeneity / total variability):   75.19%
H^2 (total variability / sampling variability):  4.03

Test for Heterogeneity:
Q(df = 1) = 4.0307, p-val = 0.0447

Model Results:

estimate       se     zval     pval    ci.lb    ci.ub
  5.7017   0.7694   7.4108   <.0001   4.1937   7.2096      ***

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3.20: Output presenting random effects analysis for mean difference of decayed teeth between mentally diseased and controls.

The results of Table 3.20 indicate that there is heterogeneity of limited significance between studies (Q=4.03, p=0.045) and there is statistically significant difference (mean difference=5.70, p<0.001) in DMFS measurement between people with mental disease and control (3.21). The funnel plot (Figure 3.10) does not indicate any

significant publication bias for the two studies. These studies present no variability according to the distance from the common value of mean difference.

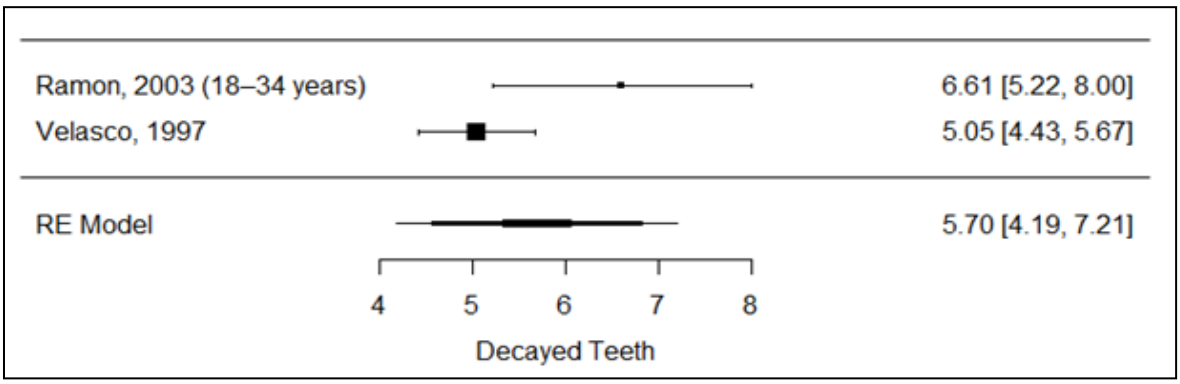

Table 3.21: Forest plot of random effects analysis for mean difference of decayed teeth between mentally diseased and controls.

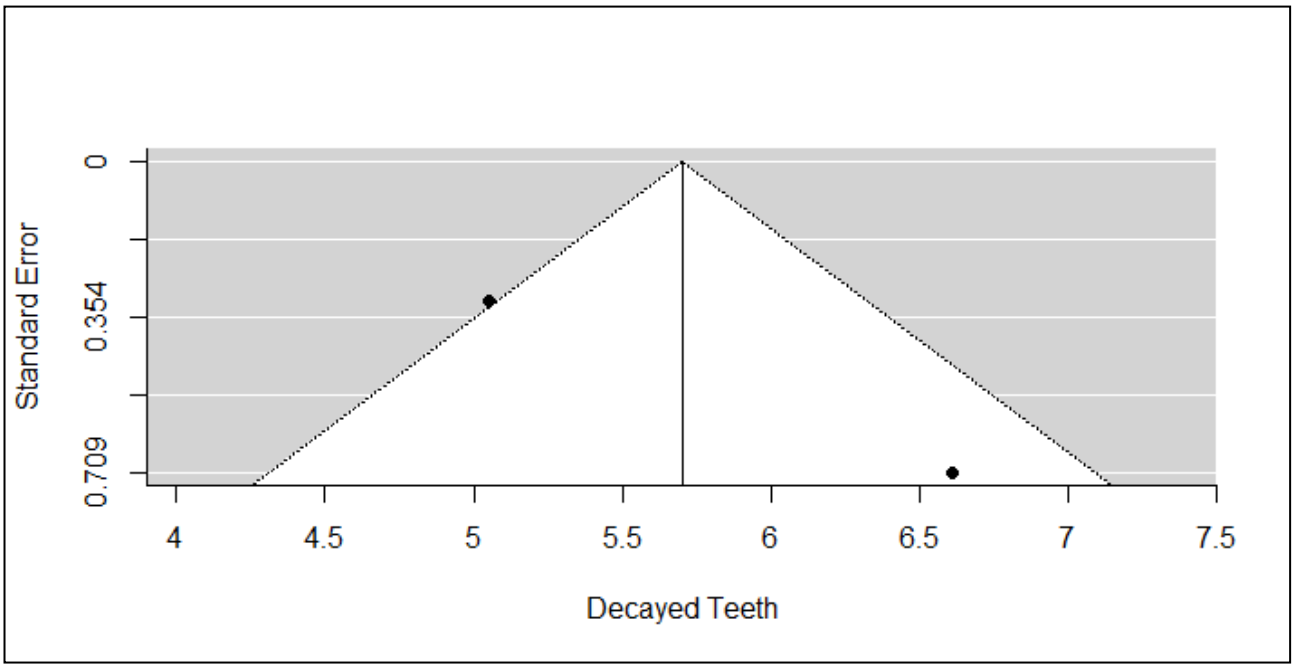

Figure 3.11: Funnel plot of mean difference of Decayed Teeth.

Next measurement under examination is the number of missing teeth. The following R commands were used for the analysis.

```
m4<-c(5.42,0.57,17.02)
mc4<-c(0.25,0.55,7.5)
sd4<-c(6.25,1.3,10.32)
sdc4<-c(0.69,1.53,6.8)
n4<-c(54,37,565)
nco4<-c(7139,29,261)
ma4<- rma(yi=m4-mc4,vi=sd4^2/n4+sdc4^2/nco4, method = "DL")
ma4
study4<-c('Ramon, 2003 (18–34 years)','Stiefel; 1990','Velasco, 1997')
forest(ma4,slab=study4,xlab='Missing Teeth')
```

```
> ma4

Random-Effects Model (k = 3; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 31.4999 (SE = 34.5703)
tau (square root of estimated tau^2 value):      5.6125
I^2 (total heterogeneity / total variability):   98.96%
H^2 (total variability / sampling variability):  96.20

Test for Heterogeneity:
Q(df = 2) = 192.3954, p-val < .0001

Model Results:

estimate       se     zval     pval    ci.lb    ci.ub
  4.8900   3.2610   1.4995   0.1337  -1.5015  11.2815

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3.22: Output presenting random effects analysis for mean difference of missing teeth between mentally diseased and controls.

The results of Table 3.22 indicate that there is highly significant heterogeneity between studies (Q=192.40, p<0.001) and there is no statistically significant difference (p=0.134) in DMFS measurement between people with mental disease and control (3.23). The funnel plot (Figure 3.11) indicates some publication bias for the

two of three studies . These studies present unusually low variability according to the distance from the common value of mean difference.

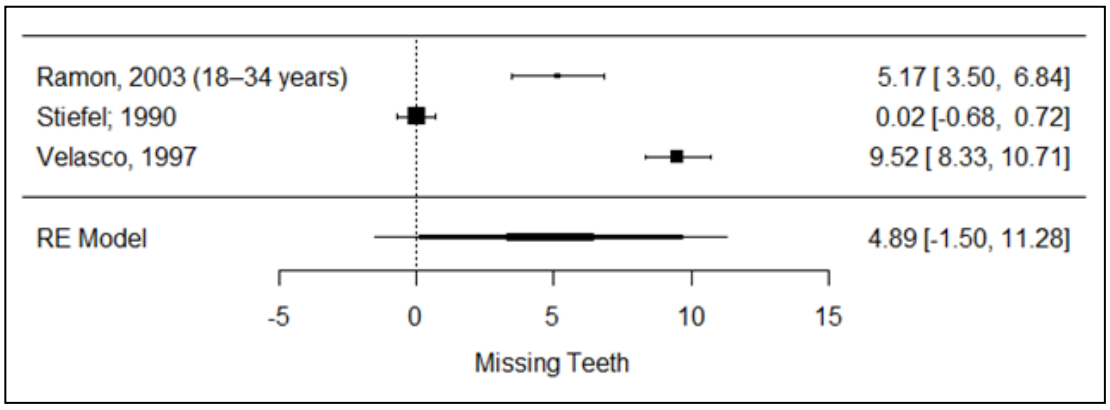

Table 3.23: Forest plot of random effects analysis for mean difference of missing teeth between mentally diseased and controls.

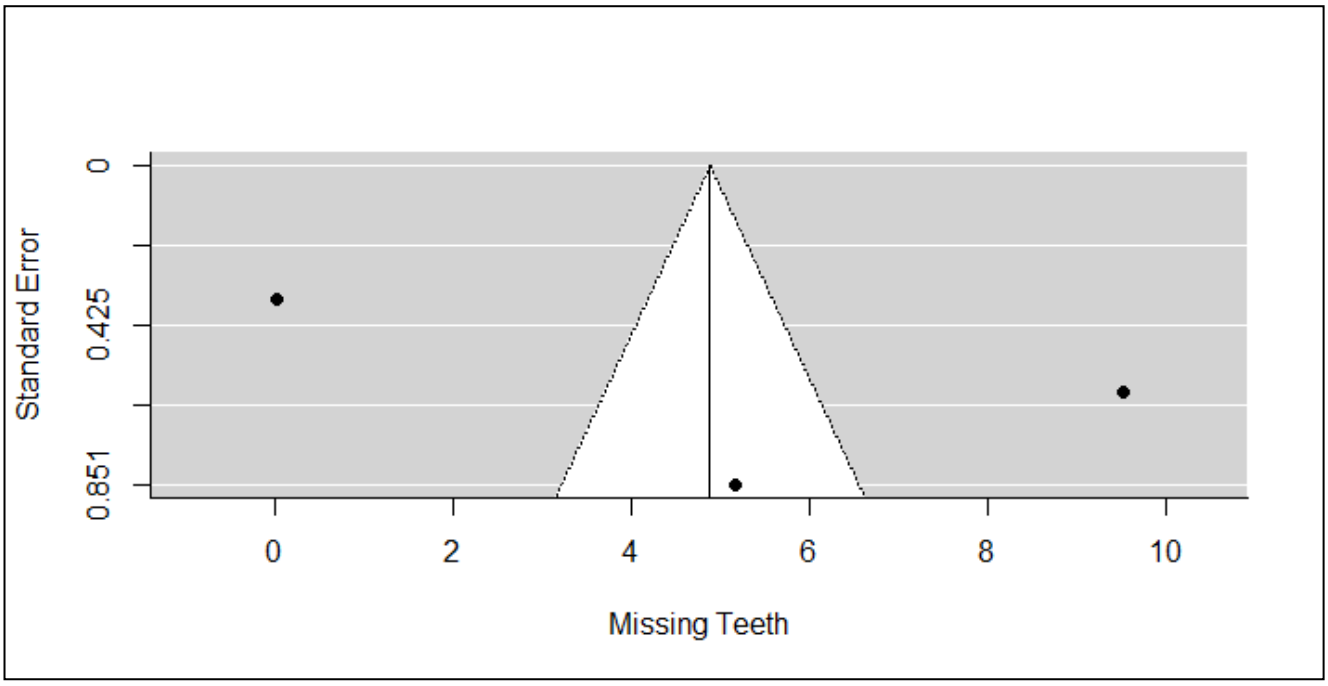

Figure 3.12: Funnel plot of mean difference of MissingTeeth.

Next measurement under examination is the number of DMFT. The following R commands were used for the analysis.

```
m5<-c(0.92,17.5,6.1,24.99)
mc5<-c(0.4,8.49,3.2,12.5)
sd5<-c(1.8,8.2,6.87,7.71)
sdc5<-c(0.92,4.95,3.49,7.1)
n5<-c(180,54,326,565)
nco5<-c(121,7139,156,261)
ma5<- rma(yi=m5-mc5,vi=sd5^2/n5+sdc5^2/nco5, method = "DL")
ma5
study5<-c('Kumar, 2006','Ramon, 2003 (18–34 years)','Rekha, 2002','Velasco, 1997')
forest(ma5,slab=study5,xlab='DMFT')
```

```
> ma5

Random-Effects Model (k = 4; tau^2 estimator: DL)

tau^2 (estimated amount of total heterogeneity): 32.6226 (SE = 33.7584)
tau (square root of estimated tau^2 value):      5.7116
I^2 (total heterogeneity / total variability):   99.39%
H^2 (total variability / sampling variability):  165.06

Test for Heterogeneity:
Q(df = 3) = 495.1874, p-val < .0001

Model Results:

estimate       se     zval     pval    ci.lb    ci.ub
  6.1965   2.8751   2.1552   0.0311   0.5614  11.8316        *

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3.24: Output presenting random effects analysis for mean difference of DMFT between mentally diseased and controls.

The results of Table 3.24 indicate that there is highly significant heterogeneity between studies (Q=495.19, p<0.001) and there is statistically significant difference (mean difference=6.20, p=0.031) in DMFS measurement between people with mental disease and control (3.25).

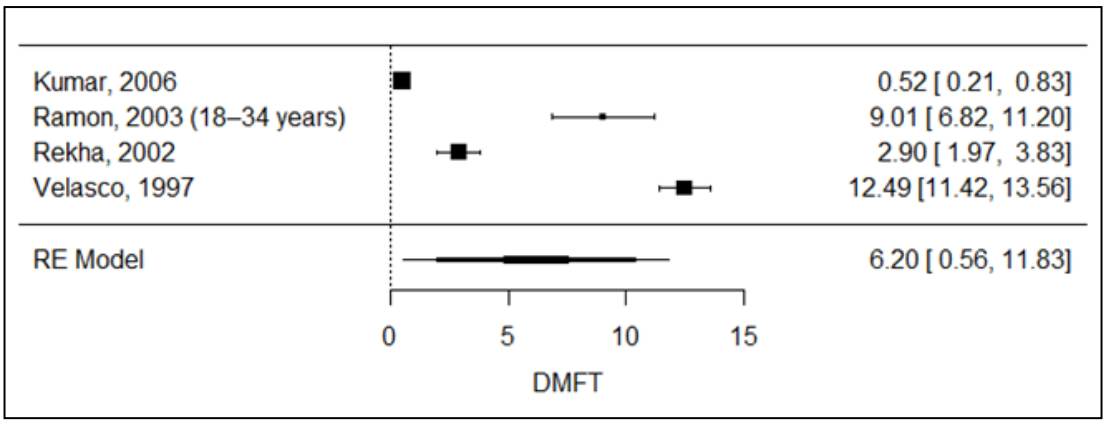

Table 3.25: Forest plot of random effects analysis for mean difference of DMFT between mentally diseased and controls.

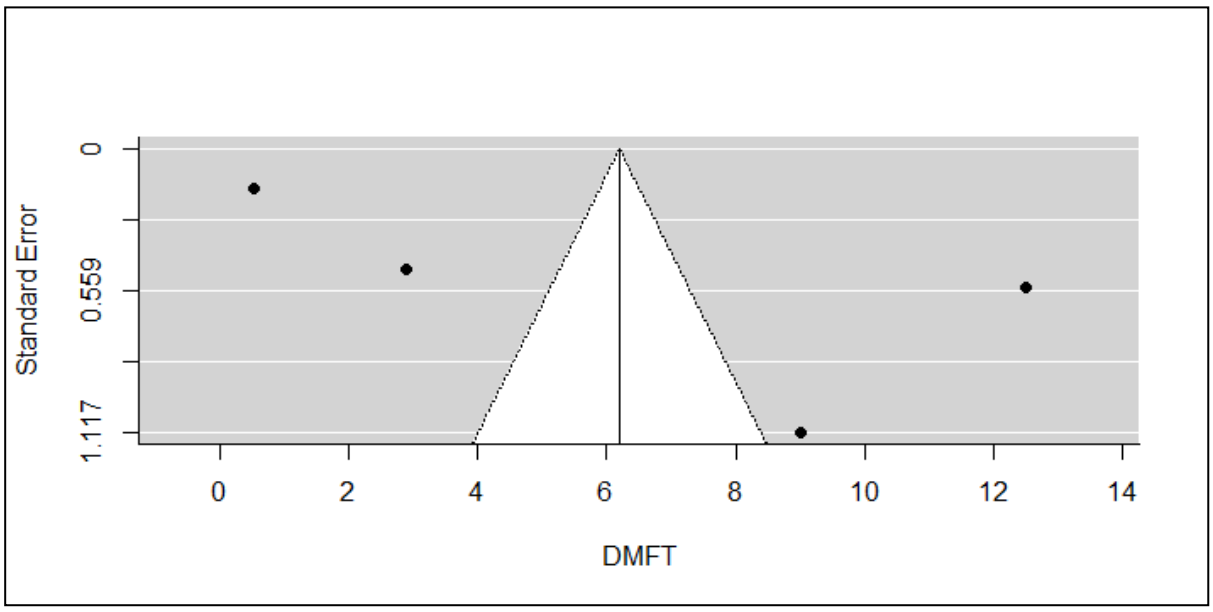

Figure 3.13: Funnel plot of mean difference of DMFT.

3.5 DISCUSSION OF RESULTS

After the analysis that was presented in this chapter we ended up to the following conclusions:

Edentulousness	**Fixed effects**	**Random effects**	τ^2	I^2 (%)
Mantel-Haenszel/ Odds ratio	3.35(<0.001)	3.25(0.002)	1.23	93.4
Mantel-Haenszel/ Risk ratio	2.09(<0.001)	2.43(0.001)	0.55	94.2
Mantel-Haenszel/ Mean difference	0.16(<0.001)	0.13(0.036)	0.04	98.0
Peto/Odds ratio	3.55(<0.001)	4.33(<0.001)	1.28	94.0
Inverse variance/ Odds ratio	2.97(<0.001)	3.34(0.002)	1.22	93.4
Inverse variance/ Risk ratio	1.65(<0.001)	2.39(<0.001)	0.48	93.4
Inverse variance/ Mean difference	0.08(<0.001)	0.13(0.017)	0.03	97.4

Table 3.26: Summary of meta-analysis of edentulousness for fixed and random effects(p-values in parenthesis) along with τ^2 estimates and heterogeneity index I^2.

- According to Table 3.26 we reach the following conclusions:
- Both Fixed and Random effects approaches for edentulousness indicated highly statistical differences between mentally diseased and control groups.
- Odds ratio estimates for edentulousness represented these differences in a clearer way in comparison to risk ratio and risk differences.
- Differences between Fixed and Random effects approaches for edentulousness were not acute except Risk ratios and Odds ratio Peto estimate.
- High heterogeneity was measured using all methods for edentulousness. This heterogeneity is also confirmed by τ^2 values.

Measurement	Mean difference	τ^2	I^2 (%)
Decayed surfaces	3,42(<0,001)	0,093	5,12
DMFS	14,60(0,007)	64,520	75,01
Decayed teeth	5,71(<0,001)	0,915	75,19
Missing teeth	4,89(0,134)	31,500	98,96
DMFT	6,20(0,031)	32,623	99,39

Table 3.27: Summary of meta-analysis of the other measurements for random effects (p-values in parenthesis) of mean differences along with τ^2 estimates and heterogeneity index I^2.

- According to Table 3.27 we reach the following conclusions:
- All differences between the two comparison groups were statistically significant apart from the "Missing teeth" measurement.
- The difference in "DMFS" measurement between mentally diseased and controls was particularly large.
- Heterogeneity was not particularly present apart from "Missing teeth" and "DMFT" measurements.

CHAPTER 4: CONCLUSION

The following discussion is based on (Kisely et all., 2011). It is well known (Lawrence et al., 2000) that individuals with severe mental illness have high rates of physical ill-health, including diabetes, cardiovascular disease, chronic lung disease and cancer. This in turn, is associated with increased mortality from preventable physical disease, so that people with schizophrenia die 15–20 years earlier than the general population. Although the oral health of the general population has improved in much of the world, psychiatric patients remain at a disadvantage in a wide range of countries. This mirrors findings in other areas such as cardiovascular disease, where the health of the general population has improved but not that of people with severe mental illness (Lawrence et al., 2003).

The results for the primary outcome, edentulousness, were strongly significant. The findings for DMFS and DMFT scores were less acute but still significant. This is possibly because both are more appropriate for dentate patients. It is impossible to record accurately the number of decayed or filled teeth if they have been lost through dental disease.

4.1 LIMITATIONS

There are some limitations to this study. According the authors (Kisely et all., 2011):

- There was considerable variation in outcome measures and how these were reported. Most studies had no comparison group and it was difficult to find suitable community controls for many of the others. Although nine studies (n = 1622) were included for the meta-analysis of the primary outcome (edentulousness), and there were fewer studies for the other outcomes.
- Most studies did not use diagnostic criteria for the psychiatric disorders of interest.
- Although age, secular trends in oral health and water fluoride levels were taken into account, other factors such as economic status or education level were difficult to be determined.

- Many of our results showed heterogeneity. This has to be further explored using sensitivity analyses of the effects of excluding outlying studies. Random effects model. Therefore, a random effects model was fit to the data when a very small number of studies was available. However, since strong heterogeneity was present the analysis results should be treated with caution.

4.2 EXPLANATIONS

Explanations for these findings (Kisely et all., 2011) include poor oral hygiene resulting in plaque formation and gingivitis. As with other aspects of physical ill-health, alcohol and substance use, tobacco and diet (including the consumption of carbonated drinks) also contribute to poor oral health. For instance, edentulousness is associated with low fruit and vegetable intake in marginalized older adults even after adjusting for sociodemographic and behavioral variables (Tsakos et al., 2010). Smoking leads to an increased incidence of erosion, cervical abrasion and gingival necrosis, and other mucosal lesions are reported in people using oral cocaine (Krutchkoff et al., 1990). Psychotropic medications can also contribute to dental disease as many cause dry mouth (xerostomia) through reduced salivary flow (Sjogren et al., 2000). Relevant medications include conventional and atypical antipsychotics, all classes of antidepressants, and mood stabilizers (Sreebny et al., 1997). Xerostomia has been found to decrease overall quality of life (Thomson et al., 2006), increase plaque and calculus formation, and lead to a higher incidence of caries, gingivitis and periodontitis (Cormac et al., 1999). People with severe mental illness may also have priorities other than their oral health, or lack privacy for oral hygiene owing to poor housing or homelessness. These issues are compounded by difficulties with access to dental care. People with severe mental illness may be reluctant to seek treatment because of the fear of pain or dental phobia, possibly exacerbated by the cost of dental care. With severe tooth loss, some measures of caries such as the number of decayed teeth actually fall. To this must be added the effects of societal and cultural differences between countries. Further research is needed to clarify how all these factors contribute to differences in findings between studies. In terms of protective factors, the presence of fluoride in the water supply should benefit all sectors of the population including those with severe mental illness.

APENDIX (R CODE)

```
data_psych <-
data.frame(total=c(135,220,138,84,109,83,326,29,91,407),
                   events=c(89,16,14,23,2,18,205,2,6,256))
data_control <-
data.frame(total=c(219,2667,3630,261,5759,8592,188,302,375,455),
                   events=c(109,213,200,8,58,1352,94,3,0,43))

study <-
c('Adam,2006','Burchell,2006','Chalmers,1998','Hede,1992','Hede,1995(
35-49 years)','Hede,1995(65-78
years)','Lewis,2001','Mirza,2001','Tang,2004','Viglid,1993')
data_sbg1<-
data.frame(row.names=study,Psychiatric=data_psych,Control=data_con
trol)
data_sbg1
studyf <- c('1','2','3','4','5','6','7','8','9','10')

library(meta)
mOR<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_
sbg1$Control.events,data_sbg1$Control.total,sm="OR" , method ="MH"
, MH.exact = TRUE)
mOR
forest(mOR,studlab=study)
funnel(mOR,studlab=studyf)
```

```
mRR<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_
sbg1$Control.events,data_sbg1$Control.total,sm="RR" , method ="MH"
, MH.exact = TRUE)
mRR
forest(mRR,studlab=study)
funnel(mRR,studlab=studyf)

mRD<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_
sbg1$Control.events,data_sbg1$Control.total,sm="RD" , method ="MH"
, MH.exact = TRUE)
mRD
forest(mRD,studlab=study)
funnel(mRD,studlab=studyf)

pOR<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_
sbg1$Control.events,data_sbg1$Control.total,sm="OR" , method
="Peto" , MH.exact = TRUE)
pOR
forest(pOR,studlab=study)
funnel(pOR,studlab=studyf)

ivOR<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_
sbg1$Control.events,data_sbg1$Control.total,sm="OR" , method
="Inverse" , MH.exact = TRUE)
ivOR
```

```
forest(ivOR,studlab=study)
funnel(ivOR,studlab=studyf)

ivRR<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_
sbg1$Control.events,data_sbg1$Control.total,sm="RR" , method
="Inverse" , MH.exact = TRUE)
ivRR
forest(ivRR,studlab=study)
funnel(ivRR,studlab=studyf)

ivRD<-
metabin(data_sbg1$Psychiatric.events,data_sbg1$Psychiatric.total,data_
sbg1$Control.events,data_sbg1$Control.total,sm="RD" , method
="Inverse" , MH.exact = TRUE)
ivRD
forest(ivRD,studlab=study)
funnel(ivRD,studlab=studyf)

library(metafor)
m1<-c(3.1,5.9,6.4)
mc1<-c(0.9,1.5,2.07)
sd1<-c(11.6,8.3,6.67)
n1<-c(109,23,37)
sdc1<-c(0.1,0.3,2.53)
nco1<-c(762,353,29)
ma1<- rma(yi=m1-mc1,vi=sd1^2/n1+sdc1^2/nco1,method = "DL")
ma1
```

```
study1<-c('Hede, 1995(35–49 years)','Hede, 1995 (65–78 years)','Stoefe;
1990')
forest(ma1,slab=study1,xlab='Subgroup of Decayed surfaces')


m2<-c(68.3,120.2,31.9)
mc2<-c(46.6,104.1,27.4)
sd2<-c(33.3,27.8,22)
sdc2<-c(0.7,1.7,20)
ma2<- rma(yi=m2-mc2,vi=sd2^2/n1+sdc2^2/nco1,method = "DL")
ma2
study2<-c('Hede, 1995(35–49 years)','Hede, 1995 (65–78 years)','Stiefel;
1990')
forest(ma2,slab=study2,xlab='DMFS')

m3<-c(9.16,7.95)
mc3<-c(2.55,2.9)
sd3<-c(5.2,6.86)
sdc3<-c(3.12,2.19)
n3<-c(54,565)
nco3<-c(7139,261)
ma3<- rma(yi=m3-mc3,vi=sd3^2/n3+sdc3^2/nco3, method = "DL")
ma3
study3<-c('Ramon, 2003 (18–34 years)','Velasco, 1997')
forest(ma3,slab=study3,xlab='Decayed Teeth')

m4<-c(5.42,0.57,17.02)
mc4<-c(0.25,0.55,7.5)
sd4<-c(6.25,1.3,10.32)
```

```
sdc4<-c(0.69,1.53,6.8)
n4<-c(54,37,565)
nco4<-c(7139,29,261)
ma4<- rma(yi=m4-mc4,vi=sd4^2/n4+sdc4^2/nco4, method = "DL")
ma4
study4<-c('Ramon, 2003 (18–34 years)','Stiefel; 1990','Velasco, 1997')
forest(ma4,slab=study4,xlab='Missing Teeth')

m5<-c(0.92,17.5,6.1,24.99)
mc5<-c(0.4,8.49,3.2,12.5)
sd5<-c(1.8,8.2,6.87,7.71)
sdc5<-c(0.92,4.95,3.49,7.1)
n5<-c(180,54,326,565)
nco5<-c(121,7139,156,261)
ma5<- rma(yi=m5-mc5,vi=sd5^2/n5+sdc5^2/nco5, method = "DL")
ma5
study5<-c('Kumar, 2006','Ramon, 2003 (18–34 years)','Rekha,
2002','Velasco, 1997')
forest(ma5,slab=study5,xlab='DMFT')
```

REFERENCES

Borenstein, M. L. V., Hedges, Higgins, J. P. T. and Rothstein, H.R. (2009). Introduction to Meta-Analysis, Wiley.

Whitehead, A. (2002). Meta-Analysis of Controlled Clinical Trials, Wiley.

Viechtbauer, V. (2010). Conducting Meta-Analyses in R with the metafor Package, Journal of Statistical Software, 36(3).

Egger, M., Davey-Smith, Altman, D.G. () Systematic Reviews in Health Care: Meta-Analysis in Context,

https:/cran.r-project.org/web/packages/meta

https:/cran.r-project.org/web/packages/metaphor

Egger, M. Smith, G.D., Schneider, M. and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ.

Sedgwick, P. (2013). Meta-analyses: how to read a funnel plot. BMJ 2013;346:f1342 doi: 10.1136/bmj.f1342.

Lewis, S. and Clarke, M. (2001). Forest plots: trying to see the wood and the trees, British Medical Journal.

Centre for evidence-based intervention, University of Oxford

Sterne et. al.(2011). Research Methods & Reporting Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ.

Terrin N, Schmid CH, Lau J.(2005). In an empirical evaluation of the funnel plot, researchers could not visually identify publication bias. Journal of Clinical Epidemiology,58:894-901.

Lawrence D, Jablensky AV, Holman CD, Pinder TJ.(2000). Mortality in Western Australian psychiatric patients. Soc Psychiatry Psychiatr Epidemiol.

Lawrence DM, Holman CDJ, Jablensky AV, Hobbs MST.(2003). Death rate from ischaemic heart disease in Western Australian psychiatric patients. Br J Psychiatry.

Mirza I, Day R, Wulff-Cochrane V, Phelan M.(2001). Oral health of psychiatric in-patients. A point prevalence survey of an inner-city hospital. Psychiatr Bull, 25: 143–5.

Cullinan MP, Ford PJ, Seymour GJ.(2009). Periodontal disease and systemic health: current status. Aust Dent J, 54 (suppl 1): S62–9.

Chapple IL.(2009). The impact of oral disease upon systemic health – symposium overview. J Dent, 37: S568–71.

Haumschild MS, Haumschild RJ.(2009) The importance of oral health in long-term care. J Am Med Dir Assoc,10: 667–71.

Williams RC, Barnett AH, Claffey N, Davis M, Gadsby R, Kellett M, et al.(2008) The potential impact of periodontal disease on general health: a consensus view. Curr Med Res Opin , 24: 1635–43.

Humphrey LL, Fu R, Buckley DI, Freeman M, Helfand M.(2008) Periodontal disease and coronary heart disease incidence: a systematic review and meta-analysis. J Gen Intern Med, 23: 2079–86.

Desvarieux M, Demmer RT, Rundek T, Boden-Albala B, Jacobs DR, Papapanou PN, et al.(2003). Relationship between periodontal disease, tooth loss, and carotid artery plaque: the Oral Infections and Vascular Disease Epidemiology Study (INVEST). Stroke,34: 2120–5.

Shultis WA, Weil EJ, Looker HC, Curtis JM, Shlossman M, Genco RJ, et al.(2007). Effect of periodontitis on overt nephropathy and end-stage renal disease in type 2 diabetes. Diabetes Care, 30: 306–11.

Azarpazhooh A, Leake JL (2006). Systematic review of the association between respiratory diseases and oral health. J Periodontol ,77: 1465–82.

Bardow A, Nyvad B, Nauntofte B.(2001). Relationships between medication intake, complaints of dry mouth, salivary flow rate and composition, and the rate of tooth demineralization in situ. Arch Oral Biol , 46: 413–23.

Lewis S, Jagger RG, Treasure E.(2001). The oral health of psychiatric in-patients in South Wales. Spec Care Dentist, 21: 182–6.

Ramon T, Grinshpoon A, Zusman SP, Weizman A.(2003). Oral health and treatment needs of institutionalized chronic psychiatric patients in Israel. Eur Psychiatry, 18: 101–5.

Selwitz RH, Ismail AI, Pitts NB.(2007). Dental caries. Lancet , 369: 51–9.

Pihlstrom BL, Michalowicz BS, Johnson NW.(2005). Periodontal diseases. Periodontal diseases. Lancet, 366: 1809–20.

Cormac I, Jenkins P.(1999). Understanding the importance of oral health in psychiatric patients. Adv Psychiatr Treat , 5: 53–60.

Roberts-Thomson K, Do L.(2007). Oral health status. In Australia's Dental Generations: The National Survey of Adult Oral Health 2004–06(eds GD Slade, AJ Spencer, KF Roberts-Thomson): 81–137. Cat. no. DEN 165. Dental Statistics and Research Series no. 34. Australian Institute of Health and Welfare.

Rekha R, Hiremath SS, Bharath S. (2002). Oral health status and treatment requirements of hospitalized psychiatric patients in Bangalore city: a comparative study. J Indian Soc Pedod Prev Dent, 20: 63–7.

Mandal KP, Tewari AB, Chawla HS, Gauba KD. (2001). Prevalence and severity of dental caries and treatment needs among population in the Eastern states of India. J Indian Soc Pedod Prev Dent, 19: 85–91.

AIHW Dental Statistics and Research Unit. The National Survey of Adult Oral Health 2004–06: Victoria. Cat. no. DEN 181. Dental Statistics and Research Series no. 45. Australian Institute of Health and Welfare, 2008.

Petersen PE, Kjoller M, Christensen LB, Krustrup U.(2004). Changing dentate status of adults, use of dental health services, and achievement of national dental health goals in Denmark by the year 2000. J Public Health Dent, 64:127–35.

Kelly M, Steele J, Nuttall N, Bradnock G, Morris J, Nunn J, et al.(2000). Adult Dental Health Survey: Oral Health in the United Kingdom 1998. TSO (The Stationery Office).

Marthaler TM, O'Mullane DM, Vrbic V.(1996). The prevalence of dental caries in Europe 1990–1995.. ORCA Saturday afternoon symposium 1995. Caries Res, 30: 237–55.

American Dental Association/Centers for Disease Control and Prevention (2006). Water Fluoridation: Nature's Way to Prevent Tooth Decay. ADA/CDC, (http://www.cdc.gov/fluoridation/pdf/natures_way.pdf).

Stillman-Lowe C (ed) (2004). One in a Million: The Facts About Water Fluoridation. British Fluoridation Society, UK Public Health Association, British Dental Association and Faculty of Public Health.

Higgins JPT, Green S (eds) (2009). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2. Cochrane Collaboration, (http:// www.cochrane-handbook.org).

Higgins JP, Thompson SG, Deeks JJ, Altman DG.(2003). Measuring inconsistency in meta-analyses. BMJ , 327: 557–60.

Stiefel DJ, Truelove EL, Menard TW, Anderson VK, Doyle PE, Mandel LS.(1990). A comparison of the oral health of persons with and without chronic mental illness in community settings. Spec Care Dentist , 10: 6–12.

Hede B, Peterson PE.(1992). Self-assessment of dental health among Danish noninstitutionalized psychiatric patients. Spec Care Dentist , 12: 33–6.

Vigild M, Brinck JJ, Christensen J.(1993). Oral health and treatment needs among patients in psychiatric institutions for the elderly. Community Dent Oral Epidemiol , 21: 169–71.

Hede B.(1995). Oral health in Danish hospitalized psychiatric patients. Community Dent Oral Epidemiol , 23: 44–8.

Angelillo IF, Nobile CG, Pavia M, De Fazio P, Puca M, Amati A.(1995). Dental health and treatment needs in institutionalized psychiatric patients in Italy. Community Dent Oral Epidemiol , 23: 360–4.

Thomas A, Lavrentzou E, Karouzos C, Kontis C. (1996). Factors which influence the oral condition of chronic schizophrenia patients. Spec Care Dentist , 16: 84–6.

Velasco E, Machuca G, Martinez-Sahuquillo A, Rios V, Lacalle J, Bullon P(1997). Dental health among institutionalized psychiatric patients in Spain. Spec Care Dentist , 17: 203–6.

Chalmers JM, Smith KD, Carter K.(1998). A multidisciplinary dental program for community-living adults with chronic mental illness. Spec Care Dentist ,18: 194–201.

Tang WK, Sun FC, Ungvari GS, O'Donnell D.(2004). Oral health of psychiatric in-patients in Hong Kong. Int J Soc Psychiatry, 50: 186–91.

Kumar M, Chandu GN, Shafiulla MD.(2006). Oral health status and treatment needs in institutionalized psychiatric patients: one year descriptive cross sectional study. Indian J Dent Res, 17: 171–7.

Adam H, Preston AJ. (2006). The oral health of individuals with dementia in nursing homes. Gerodontology, 23: 99–105.

Burchell A, Fembacher S, Lewis R, Neil A.(2006). 'Dental as anything': inner south community health service dental outreach to people with a mental illness. Austr J Primary Health, 12: 75–82.

Bhansali S, Tripathi A, Tiwari SC, Singh SV. (2008). A study of the prosthodontic and oral health needs of an ageing psychiatric population. Gerodontology, 25: 113–7.

Persson K, Axtelius B, Soderfeldt B, Ostman M.(2009). Monitoring oral health and dental attendance in an outpatient psychiatric population. J Psychiatr Ment Health Nurs, 16: 263–71.

Ponizovsky AM, Zusman SP, Dekel D, Masarwa AE, Ramon T, Natapov L, et al (2009). Effect of implementing dental services in Israeli psychiatric hospitals on the oral and dental health of inpatients. Psychiatr Serv, 60: 799–803.

Rudolph MJ, Chikte UM.(1993). Dental caries experience and peridontal disease in institutionalised male psychiatric patients. J Dent Assoc S Africa, 48: 451–4.

Zusman SP, Ponizovsky AM, Dekel D, Masarwa AE, Ramon T, Natapov L, et al.(2010). An assessment of the dental health of chronic institutionalized patients with psychiatric disease in Israel. Spec Care Dentist, 30: 18–22.

Kenkre AM, Spadigam AE.(2000). Oral health and treatment needs in institutionalized psychiatric patients in India. Indian J Dent Res, 11: 5–11.

AIHW Dental Statistics and Research Unit. The National Survey of Adult Oral Health 2004–06: New South Wales. Cat. no. DEN 176. Dental Statistics and Research Series no. 40. Australian Institute of Health and Welfare.

Kirkegaard E, Borgnakke WS, Gronbrek L(1986). Oral health status, dental treatment need, and dental care habits in a representative sample of the adult Danish population. . Royal Dental College.

Oral Health Education Unit. 2001 Oral Health Survey. Department of Health of the Government of the Hong Kong Special Administrative Region, 2006 (http://www.toothclub.gov.hk/en/en_home_06_01.htm).

Palmqvist S, Soderfeldt B, Vigild M, Kihl J.(2000). Dental conditions in middle-aged and older people in Denmark and Sweden: a comparative study of the influence of socioeconomic and attitudinal factors. Acta Odontol Scand, 58: 113–8.

Sgan-Cohen HD, Katz J, Horev T, Dinte A, Eldad A.(2000). Trends in caries and associated variables among young Israeli adults over 5 decades. Community Dent Oral Epidemiol, 28: 234–40.

Alvarez-Arenal A, Alvarez-Riesgo JA, Petia Lopez JM, Fernandez Vazquez IP Villa Vigil MA.(1996). DMFT and treatment needs in adult population of Oviedo, Spain. Spain. Community Dent Oral Epidemiol, 24: 17–20.

Krustrup U, Petersen PE.(2007) Dental caries prevalence among adults in Denmark – the impact of socio-demographic factors and use of oral health services. Community Dent Health, 24: 225–32.

Lawrence D, Jablensky AV, Holman CD, Pinder TJ.(2000). Mortality in Western Australian psychiatric patients. Soc Psychiatry Psychiatr Epidemiol, 35: 341–7.

Lawrence DM, Holman CDJ, Jablensky AV, Hobbs MST.(2003). Death rate from ischaemic heart disease in Western Australian psychiatric patients 1980–1998. Br J Psychiatry, 182: 31–6.

Office for National Statistics. Adult Dental Health Survey. Oral Health in the United Kingdom 1998: 1–12. TSO (The Stationery Office), 2000.

Tsakos G, Herrick K, Sheiham A, Watt RG (2010). Edentulism and fruit and vegetable intake in low-income adults. J Dent Res, 89: 462–7.

Krutchkoff DJ, Eisenberg E, O'Brien JE, Ponzillo JJ.(1990) Cocaine-induced dental erosions. N Engl J Med, 322: 408.

Sjogren R, Nordstrom G.(2010). Oral health status of psychiatric patients. J Clin Nurs, 9: 632–8.

Sreebny LM, Schwartz SS. (1997). A reference guide to drugs and dry mouth – 2nd edition. Gerodontology, 14: 33–47.

Thomson WM, Lawrence HP, Broadbent JM, Poulton R.(2006). The impact of xerostomia on oral-health-related quality of life among younger adults. Health Qual Life Outcomes, 4: 86.

Barnes GP, Allen EH, Parker WA, Lyon TC, Armentrout W, Cole JS.(1988). Dental treatment needs among hospitalized adult mental patients. Spec Care Dentist , 8: 173–7.

Australian Research Centre for Population Oral Health, The University of Adelaide. Oral Health Promotion Clearinghouse (https://www.adelaide.edu.au/oral-health-promotion/).

Almomani F, Brown C, Williams KB.(2006). The effect of an oral health promotion program for people with psychiatric disabilities. Psychiatr Rehabil J, 29: 274–81.

General Practice Queensland. Activate: Mind and Body. General Practice Queensland,2009(http://www.gpqld.com.au/page/Programs/Mental_Health/Improving_the_Physical_Health_of_People_with_a_Severe_Mental_Illness_Project/).

Michael J. Bradburn1; ‡, Jonathan J. Deeks1; †, Jesse A. Berlin2; §and A. Russell Localio2v Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events.

Greenland, S and Salvan, A. Bias in the one-step method for pooling study results

Whitehead, A. Combining Estimates of a Treatment Difference Across Trials Meta-Analysis of Controlled Clinical Trials.

Jonathan J Deeks, Douglas G Altman, Michael J Bradburn Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis

A B Haidich.Meta-analysis in medical research.

Julian P T Higgins, **Simon G Thompson, Jonathan J Deeks and Douglas G Altman**. Measuring inconsistency in meta-analyses.


ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ